DOCUMENT RESUME

ED 448 598 FL 026 520

AUTHOR	Porter, Rosalie Pedalino, Ed.
TITLE	Testing in Texas: Accountability for Bilingual Students. READ Perspectives, Fall 2000.
INSTITUTION	READ: Research in English Acquisition and Development Inst., Inc., Washington, DC.; Center for Equal Opportunity, Washington, DC.
ISSN	ISSN-1091-6822
PUB DATE	2000-00-00
NOTE	159p.; For the 1999 issue (Vol. 6), see ED 439 611. Published annually.
AVAILABLE FROM	Read Institute, Center for Equal Opportunity, 815 15th Street, NW, Suite 928, Washington, DC 20005. Tel: 202-639-0803; Fax: 202-639-0827; Web site: http://www.ceousa.org.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Read Perspectives; v7 Fall 2000
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Bilingual Education Programs; Elementary Secondary Education; *English (Second Language); Language Minorities; *Limited English Speaking; Literacy Education; Politics of Education; Second Language Instruction; Second Language Learning; Spanish Speaking; *Student Evaluation; *Testing
IDENTIFIERS	*Texas

ABSTRACT

This issue has a special, multi-article section on student testing in Texas and contains three additional and unrelated articles. "The Texas Testing Case Documents: G.I. Forum, et al. v. Texas Education Agency, et al." section has five articles: "Overview" (Roger Clegg); a copy of the "First Amended Complaint"; "Expert Reports" (Susan E. Phillips, William A. Mehrens, Rosalie P. Porter); a copy of the "Final Judgment and Order"; and a postscript, "Testing the Academic Achievements of Limited English Proficient Students" (Rosalie P. Porter). Other articles include the following: "Recognizing Successful Schools for High Achieving, Low-Income Students: The 'No Excuses' Campaign" (Robert E. Rossier); "Bilingual Students and MCAS: Some Bright Spots in the Gloom" (Ralph E. Beals, Rosalie P. Porter); and "Different Questions, Different Answers: A Critique of the Hakuta, Butler, and Witt Report" (Christine H. Rossell). Tables, figures, and references are included in each article. (KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 448 598

READ PERSPECTIVES

FALL 2000 ■ Vol. VII

Testing in Texas

Accountability for Bilingual Students

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Rosalie Porter

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Also Inside:

Beals & Porter on Massachusetts Testing
Robert Rossier Offers 'No Excuses'
Christine Rossell Critiques Hakuta et al.

FL 026 520

A PUBLICATION OF THE READ INSTITUTE
& THE CENTER FOR EQUAL OPPORTUNITY

READ
PERSPECTIVES

FALL 2000 ■ Vol. VII

Contents

Introduction—*Rosalie P. Porter* **Page 3**

The Texas Testing Case Documents:
G.I. Forum, et al. v. Texas Education Agency, et al. **Page 7**

Overview—*Roger Clegg* / Page 7

First Amended Complaint / Page 15

Expert Reports—*Susan E. Phillips* / Page 23
—*William A. Mehrens* / Page 59
—*Rosalie P. Porter* / Page 73

Final Judgment and Order / Page 79

Postscript: Testing the Academic Achievement of Limited-English Proficient Students—*Rosalie P. Porter* / Page 102

Recognizing Successful Schools for High Achieving, Low-Income Students: The "No Excuses" Campaign **Page 110**
—*Robert E. Rossier*

Bilingual Students and the MCAS: Some Bright Spots in the Gloom **Page 117**
—*Ralph E. Beals and Rosalie P. Porter*

Different Questions, Different Answers: A Critique of the Hakuta, Butler, and Witt Report **Page 134**
—*Christine H. Rossell*

About the Authors **Page 155**

3

READ PERSPECTIVES

A publication of the READ Institute

EDITOR
Rosalie Pedalino Porter, Ed.D., *Editor*

ACADEMIC REVIEW PANEL
Ralph E. Beals, Ph.D.
Professor of Economics
Amherst College

Jim Littlejohn
Former Program Director
Office of Civil Rights
U.S. Department of Education

Sandra Stotsky, Ed.D.
Deputy Commissioner
Massachusetts Department
of Education

Kevin Clark
Clark Consulting Group
Stockton, Calif.

Robert E. Rossier, Ph.D.
Educational Consultant
Los Angeles, Calif.

Carol Whitten
Former Director
Office of Bilingual Education and
Minority Languages Affairs
U.S. Department of Education

Additional Readers
Norma C. McKenna, David T. Porter, Christine H. Rossell

READ Institute Staff
Linda Chavez, Litt.D. (Honor), President
Jorge Amselle, Executive Director

The READ Institute is a project of the
Center for Equal Opportunity
815 15th Street, N.W., Suite 928
Washington, DC 20005

Phone (202) 639-0803
Fax (202) 639-0827
http://www.ceousa.org

© Copyright 2000, The Center for Equal Opportunity

The Center for Equal Opportunity, founded in 1995, is a non profit, tax exempt, Section 501(c)(3) organization. CEO is a nonpartisan research and educational institution studying issues of race, ethnicity, and assimilation. This publication is one of several produced by the Center For Equal Opportunity.

Introduction

R*EAD Perspectives* announced the unifying theme for Volume VII, the present issue, to be "Accountability for Bilingual Students." This theme is addressed and broadened to encompass the larger population of minority students in U.S. public schools, due in part to the coincidence of a federal court ruling that was handed down in Texas in early 2000.

The major portion of the current magazine is devoted to the Texas lawsuit challenging the right of the state to require all students to pass a 10th grade test of basic skills in reading, writing and mathematics in order to receive a high school diploma, *G.I. Forum, et al. v. Texas Education Agency, et al.* The Mexican American Legal Defense and Educational Fund (MALDEF) brought suit to excuse black and Hispanic high school students from having to pass the 10th-grade test in order to graduate, on the grounds that minority students are not provided an equal education; that it is a denial of their civil rights to keep them from graduating from high school on the basis of one test score; because of a prior history of segregation and discrimination; and on the basis of disparate impact, i.e., minority students fail the test in higher numbers than their proportions in the school population.

Judge Edward C. Prado, who conducted the five-week court hearings, ruled in January 2000 that (1) the state does not discriminate unfairly, (2) it has provided additional resources for schools with underperforming students, and (3) the greater efforts of recent years have resulted in better performance by minority students (including Limited-English Proficient students) who are passing the 10th-grade test in greater numbers every year and graduating from high school. (Prado, 1/7/2000) One of the key elements that carried weight in the Texas deliberations is the fact that high school students are offered remedial classes and tutoring and eight opportunities to retake the 10th-grade test. The gap between passing scores for blacks, Hispanics and whites has narrowed substantially. In a *Wall Street Journal* editorial, Jay Greene of the Manhattan Institute in New York City chronicles the educational transformation in Texas:

> Through mandatory, statewide testing of public school students in reading, writing and math, Gov. Bush has been able to hold public schools accountable for results. "In 1994," writes Mr. Greene, "only 53 percent of public school students passed the [statewide test]. In 1998, 78 percent did—a

> remarkable improvement. The pass rate for blacks and Hispanics more than doubled, to 63 percent in 1998 from 31 percent in 1994. Hispanics' rate shot up to 70 percent from 39 percent." By 1998, Texas's black public school students ranked first in the country among minority students." (7/31/2000, p. A-22)

Judge Prado's ruling is of immense importance to the other states that also require a test for high school graduation and whose state policies would be at risk of being overturned. But the decision is of even greater importance in the efforts to improve schooling for minority students. Excusing minority students from being evaluated by a uniform, objective measure of basic learning—and the Texas 10th-grade test is not a rigorous test—sends the damaging message that we do not expect minorities to meet these standards and therefore neither they nor their schools can be held accountable. In February 2000, MALDEF announced that it would not appeal the ruling in *G.I. Forum*. (*Washington Post*, 2/8/2000, p. 9)

READ Perspectives provides the major documents in the *G.I. Forum* case for the interest of state education departments, school board members, school administrators and attorneys for school districts across the country. Roger Clegg, counsel for the Center for Equal Opportunity in Washington, D.C., and a specialist in civil rights issues, introduces the Texas section with an incisive review of the case. A postscript to the documents focusing on the necessity of including Limited-English Proficient (LEP) students in state testing, was written by editor Rosalie P. Porter, published first in *Applied Measurements in Education* (September 2000) and reprinted here.

Robert E. Rossier, California specialist in bilingual education issues who has contributed earlier articles to *READ Perspectives*, reviews the "No Excuses Study" published by the Heritage Foundation in Washington, D.C. (Carter, 1999). The Foundation awarded its 1999 Salvatori Prize for American Citizenship to seven school principals in schools serving mostly minority students from families of poverty, schools where there is a record of high academic achievement. The demographics of each school are detailed, and the particular values and priorities of each principal are explained. Rossier gives special attention to the Bennett-Kew Elementary School in Inglewood, Calif., whose principal, Nancy Ichinaga, has a remarkable record of success specifically for the achievement of bilingual students. The new focus on finding schools that demonstrate the academic success of a large proportion of their children from low-income homes, instead of making excuses for their academic failure, is a welcome and growing research effort that will continue to be examined in *READ Perspectives*.

The editor of *READ Perspectives* and Professor Ralph E. Beals of Amherst College have been engaged in ongoing reviews of the state testing of Limited-English Proficient students in Massachusetts for the past three years on behalf of the READ Institute. Massachusetts was first in the U.S. to legislate mandatory transitional bilingual education programs (1971). However, the state has not published any data on LEP student achievement until the recent advent of the Massachusetts Comprehensive Assessment System (MCAS), the 1993 education reform initiative that mandates testing of all students in grades 4, 8 and 10, starting in 1998.

The Beals-Porter study reviews the participation and performance of LEP students on the English Language Arts, Mathematics, and the Science and Technology tests in all 32 Massachusetts districts with 10 or more LEP students; compares rates of passing test scores and district demographics; and makes preliminary determinations of which districts are demonstrating better academic performance by LEP students, especially at the fourth-grade level. The state divides the school population into "regular students," "students with disabilities" and "Limited-English Proficient Students," and it is discouraging to note that at all grade levels and on all subjects tested, LEP student performance is at the lowest levels. Unacceptable as this may be, it is clear where the challenges lie and where resources must be focused to improve the opportunities for these students.

The main conclusion of this study is that the data collection and reporting by the Massachusetts Department of Education is seriously flawed, with these major problems: (1) data are contradictory and inconsistent in regard to the numbers of students tested; (2) LEP students who were eligible to take all the MCAS tests in English in 1999 either did not take the math and science tests in half the districts surveyed or else their test scores were not recorded; and (3) for LEP students who have been in U.S. schools lower than three years and who are literate in Spanish, the math and science tests may be taken in a bilingual (Spanish/English) version of the test, but the state Department of Education did not mark the test forms to identify who took the test in English or in the Spanish/English version. The Beals-Porter study is a first step in the state's long-neglected responsibility to account for the academic progress of LEP students.

California education policy for language minority, limited-English children is attacked by Professor Kenji Hakuta and colleagues Yuko Goto Butler and Daria Witt, all of Stanford University, and defended by Professor Christine H. Rossell of Boston University. The Hakuta paper, "How Long Does It Take English Learners To Attain Proficiency?" concludes that it takes three to five years to develop oral language fluency, and academic English proficiency can take four to seven years, based on an examination of results from four school districts, two in California and two in Canada. (Hakuta) The authors claim that the California policy of one year or so of English Immersion programs is "wildly unrealistic."

Professor Rossell offers an unambiguously negative critique of the Hakuta report. She states at the very beginning of her essay, "Different

Questions, Different Answers," that "The authors are simply wrong in believing that knowing how long it takes an LEP child to achieve parity with native English speakers or to be classified 'proficient' on an English proficiency test tells us how long they need special education services or how long they should be in a sheltered immersion classroom." In her concluding paragraph, Rossell comes down solidly in favor of Proposition 227, the "English for the Children" initiative passed by California voters in 1998, which sets a time period for LEP students to be placed in separate, below-grade level classrooms, "...not because anyone thinks non-English speaking children will have mastered English in one year, but because what evidence there is suggests that sometime during their first year, immigrant children will understand enough English so that they will be better off in a grade-level mainstream classroom than in a remedial classroom. Furthermore, if a time limit were not specified in the legislation, more than half of them would never be mainstreamed, no matter how fluent they were in English."

The seemingly never-ending debate on the rate of second-language acquisition is joined once more in the Rossell critique of the Hakuta paper, the most current round of arguments in this arena, a fitting conclusion for this volume of *READ Perspectives.*

—Rosalie Pedalino Porter, Editor

References

Carter, S. C. (1999). *No excuses: Seven principals of low-income schools who set the standards for high achievement.* Washington, D.C.: The Heritage Foundation.

Hakuta, K., Butler, Y. G., Witt, D. (January 2000). How long does it take English learners to attain proficiency? Ventura, Calif.: *University of California Linguistics Minority Research Institute.*

MALDEF announces that it will not be appealing Judge Prado's ruling. (Feb. 8, 2000). *Washington Post* (from news service), p. 9.

Prado, E. C. (Jan. 7, 2000). *Judgment and Order in G.I. Forum et al. v. Texas Education Agency et al.* Civil Action No. SA-97-CA-1278-EP, U.S. District Court, Western District of Texas, San Antonio Division.

Porter, R. P. (November 2000). Accountability is overdue: Testing the academic achievement of limited-English proficient students. *Applied Measurements in Education,* 13(2), pp. 403-410.

Wall Street Journal. (July 31, 2000). The politics of learning (unsigned editorial), p. A-22.

The Texas Testing Case Documents

G.I. Forum, et al. v. Texas Education Agency, et al.

Overview

Roger Clegg

GI Forum v. Texas Education Agency: A Summary

The litigation in *GI Forum v. Texas Education Agency* is of crucial importance to those states and school districts that already have or are considering a requirement that students pass a comprehensive test before being awarded a high school diploma. Texas is one of 19 states with such a requirement.

The lawsuit in *GI Forum* was filed on October 14, 1997, in federal district court in Texas. The complaint against the state of Texas by the Mexican American Legal Defense and Educational Fund (MALDEF) alleged that the Texas Assessment of Academic Skills (TAAS) exit test for high school graduation was illegally discriminatory. The test measures proficiency in reading, writing and math. On January 7, 2000, Judge Edward C. Prado dismissed the lawsuit, ruling that TAAS neither unfairly discriminates against black and Mexican American students nor denies them their right to due process. The next month, MALDEF announced that it would not be appealing Judge Prado's ruling. (MALDEF announces, p. 9)

READ Perspectives has collected the key materials from this case and is publishing them here. In addition to the original complaint and Judge Prado's opinion, we are also including decisive testimony from three expert witnesses at the trial: Dr. S.E. Phillips, Dr. William A. Mehrens and Dr. Rosalie Pedalino Porter.

The complaint. MALDEF's complaint was filed on behalf of the GI Forum, Image de Tejas, and seven Mexican American or African American students. It named as defendants the Texas Education Agency (TEA), members of the Texas State Board of Education, and Texas Commissioner of Education Mike Moses. The complaint asserted that TAAS "denies diplomas to Mexican American and African American students at a rate significantly higher than that of Anglo students," thereby "violat[ing] a variety of United States Constitutional, statutory and regulatory provisions, as well as fundamental fairness."

The complaint alleged that "Mexican Americans and African Americans have suffered from a long and well-documented history of discrimination in Texas public schools." It asserted that the "[w]hites are almost twice as likely as Mexican Americans and African Americans to pass the TAAS," and that "TAAS is an invalid instrument for determining which students are qualified to receive diplomas" because "[m]any who score below the cutoff score could perform satisfactorily as high school graduates in college, the military and the workforce." The core of the complaint, then, was that TAAS had an illegal "disparate impact" on blacks and Mexican Americans.

MALDEF concluded that the defendants were denying "equal educational opportunities" in contravention of an earlier federal case, *United States v. Texas,* as well as violating the plaintiffs' equal protection and due process rights under the Fourteenth Amendment to the United States Constitution. In addition, MALDEF complained that defendants were illegally discriminating on the basis of race and national origin in violation of Title VI of the Civil Rights Act of 1964, the U.S. Department of Education's Title VI regulations and the Equal Educational Opportunities Act. The complaint asked the court to enjoin the state's use of TAAS until it is "properly validated" and its discriminatory effects "shown to be as minimal as any reasonably effective alternative." Finally, MALDEF sought a permanent injunction against "any standardized test as an absolute requirement for receipt of a high school diploma."

Phillips testimony. Dr. Phillips of Michigan State University testified that the TAAS exit level test "meets all relevant professional standards for test development and use." She analyzed the differential performance between black and Mexican American students, on the one hand, and white students on the other, as well as the dropout data. In addition, she carefully pointed out the flaws in the analyses of plaintiffs' three witnesses: Dr. Martin Shapiro, Dr. Walter Haney, and Mr. Mark Fassold.

The benefits that Dr. Phillips identified from TAAS's implementation included increasing the level of skills and knowledge attained by high school graduates, better remediation for unprepared students, and closing the gap between the performance of different racial and ethnic groups. She also noted that eliminating TAAS would probably not change the dropout rate appreciably or cause African American or Mexican American students to learn more, but it would make schools less accountable, remove incentives for remediation and "reduce the value of a high school diploma in Texas."

Dr. Phillips concluded that TAAS "did not create the social problems faced by minority groups but has contributed to their improvement." She said the test should be retained because "its benefits to minority students far outweigh its alleged and unproven social costs."

Mehrens testimony. Dr. Mehrens—a colleague of Dr. Phillips at Michigan State University—testified that "tests must be judged against reasonable standards" and that "TAAS has been constructed in a professionally accepted manner." TAAS tests curricular material that the state views as important for graduates to have mastered and, indeed, Dr. Mehrens concluded that without a requirement like TAAS students might graduate without having achieved what the state has deemed to be a set of minimal requirements. Students have had ample opportunity to learn the materials TAAS tests on, and providing instruction over the objectives tested by TAAS is to be applauded, not condemned. Dr. Mehrens further testified that the approach taken by Texas with TAAS will *help* disadvantaged students and will *remove* vestiges of past discrimination.

Dr. Mehrens also testified that the test is reliable and that the eight opportunities students have to take the test ensures that the possibility of not passing due to random error is almost zero (and, indeed, means that some students who shouldn't pass, will). He resolved several other technical issues—regarding validity, potential bias, adverse impact data, and the appropriate decision-making model—in TAAS's favor.

Finally, Dr. Mehrens testified that standard setting is a judgmental process. Those in authority should make this judgment, he said, and the state Board of Education had sufficient information to set the cut-off scores.

Porter testimony. The third witness whose testimony we include is Dr. Rosalie P. Porter, an expert on bilingual education and the editor of *READ Perspectives*. Dr. Porter testified that "the accountability element" is "often lacking" in "bilingual program evaluation." She discussed in particular her experiences in Massachusetts, which are illuminating.

Dr. Porter stated, "Exempting whole groups of students from statewide assessments on the expectation that they will not perform adequately is unfair to the students who are excluded, as well as to their classmates." Furthermore, "Maintaining rigorous standards and high expectations for minority students requires that periodic assessments of each student's progress be conducted and reported."

Dr. Porter concluded that the TAAS program "is a fair test of student learning," and noted that "minority students have registered consistently higher passing levels on the 10th-grade test each year since 1995, showing more rapid rates of improvement than for white non-Hispanic students.

"To suggest that students should be granted high school diplomas without demonstrating minimal knowledge and skills on a uniform measure," she continued, "is not acceptable for the current requirements of the technological/information age job market or for pursuing higher education." Dr. Porter characterized an opposing witness's complaint regarding time wasted on "teaching the test" as "a harmful exaggeration."

The court's ruling. Judge Prado decided, after "much reflection," that "the TAAS examination does not have an impermissible adverse impact on Texas's minority students and does not violate their right to the due process of law." (The plaintiffs' other claims had already been dismissed by Judge Prado in an order dated July 27, 1999.) At the end of the judge's introduction, he concluded that "the Plaintiffs failed to prove that the [challenged] policies are unconstitutional, that the adverse impact is avoidable or more significant than the concomitant *positive* impact, or that other approaches would meet the State's articulated legitimate goals." (Emphasis in the original.)

"The court has no authority to tell the state of Texas what a well-educated high-school graduate should demonstrably know at the end of 12 years of education," Judge Prado wrote. "Ultimately, resolution of this case turns not on the validity of the parties' views on education but on the state's right to pursue educational policies that it legitimately believes are in the best interest of Texas students."

Judge Prado's order made extensive findings of fact about TAAS. On the disparate-impact issue in particular, he wrote:

> The Court finds as an inescapable conclusion that in every administration of the TAAS test since October 1990, Hispanic and African American students have performed significantly worse on all three sections of the exit exam than majority students. However, the Court also finds that it is highly significant that minority students have continued to narrow the passing rate gap at a rapid rate. In addition, minority students have made gains on other measures of academic progress, such as the National Assessment of Educational Progress test. The number of minority students taking college entrance examinations has also increased.
>
> ...
>
> The Court finds that failure of the exit-level TAAS examination during the first seven administrations results in immediate remedial efforts. At the last administration, of course, failure of the exit-level TAAS examination results in failure to receive a diploma. However, the Court finds, based on evidence presented at trial, that the effect of remediation, which is usually eventual success in passing the examination and thus receipt of a high school diploma, is more profound than the steadily decreasing minority failure rate.

Judge Prado's conclusions of law addressed, first, the disparate-impact claims under the Title VI regulations and, second, the due-process claims under the Fourteenth Amendment to the U.S. Constitution. With respect to the former, he found: "While the TAAS test does adversely affect minority students in significant numbers, the TEA has demonstrated an educational necessity for the test, and the Plaintiffs have failed to identify equally effective alternatives." With respect to the latter, Judge Prado wrote:

> The TEA has provided adequate notice of the consequences of the exam and has ensured that the exam is strongly correlated to material actually taught in the classroom. In addition, the test is valid and in keeping with current educational norms. Finally, the test does not perpetuate prior educational discrimination or unfairly hold Texas minority students accountable for the failures of the State's educational system. Instead, the test seeks to identify inequities and address them. It is not for this Court to determine whether Texas has chosen the best of all possible means for achieving these goals. The system is not perfect, but the Court cannot say it is unconstitutional.

Judge Prado also noted, "The results of TAAS are used, in many cases quite effectively, to motivate not only students but schools and teachers to raise and meet educational standards."

The Fundamental Problems with Disparate-Impact Lawsuits

The rejection of MALDEF's claim in *GI Forum* is good news for anyone who cares about education or civil rights. The whole disparate-impact approach to civil rights litigation is fundamentally flawed. MALDEF's assertion—that TAAS ought to be ruled illegal because a disproportionate number of blacks and Mexican Americans fail to pass it, even though the same test was given in the same way to all students and was drawn up with no racial or ethnic animus. This claim should be rejected out of hand, as a matter of both law and policy.

Three kinds of "discrimination." There are three kinds of racial and ethnic discrimination that can be held illegal under our federal civil rights laws. The relevant statute here is Title VI of the Civil Rights Act of 1964. It reads: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance."

The first kind is holding people to different standards, depending on the color of their skin or where their ancestors came from. If you have a double standard based on race or ethnicity, everyone would agree that this is discrimination under any normal use of the term.

A second kind of discrimination that violates federal civil rights laws is when someone chooses a selection criterion because of the racial or ethnic impact it will have. For instance, if a school was told to desegregate and then suddenly decided to change its admission criteria in order to keep out blacks, that would clearly violate the law, even if the new criteria were neutral on their face.

Here is a more recent example. The U.S. Court of Appeals for the Fifth Circuit in *Hopwood v. Texas,* 78 F.3d 932 (5th Cir. 1996), *cert. denied,* 116 S. Ct. 2581 (1997), held that the state could not use racial and ethnic admissions preferences. Texas decided, in the wake of the decision, that it would no longer consider SAT scores for the top 10 percent of each high school class. It made clear that it was changing the standard in order to ensure that more blacks and Hispanics, and thus fewer whites and Asians, were admitted. In doing so, then, Texas was clearly violating the law. MALDEF, of course, made no complaint about the new Texas law.

This leaves a third kind of discrimination, namely "disparate impact." Under this approach, a selection device that is neutral on its face, and that is applied neutrally, and that was chosen with no discriminatory animus, is nonetheless presumed to be illegal if it has a disproportionate *effect* on some racial or ethnic group.

No normal person would consider a test in such circumstances to be "discrimination" under any reasonable definition of the term. The Supreme Court has made clear that Title VI itself bans only intentional discrimination—that is, only the first two kinds of discrimination discussed. See *United States v. Fordice,* 505 U.S. 717, 732 n.7 (1992), citing *Regents of the University of California v. Bakke,* 463 U.S. 265 (1978), and *Guardians Association v. Civil Service Commission of City of New York,* 463 U.S. 582 (1983). See also *Washington v. Davis,* 426 U.S. 229 (1976), and *Village of Arlington Heights v. Metropolitan Housing Development Corp.,* 429 U.S. 252 (1977).

Nonetheless, MALDEF has decided to challenge standardized tests if they have a "disparate impact." The disparate impact approach is dubious enough in employment law, where it began, and should not be extended to other areas, particularly education.

Policy objections to the disparate-impact approach. Unfortunately, there is judicial and regulatory support for applying the disparate-impact model to education, although there is a good chance that it will be rejected out of hand if it reaches the Supreme Court. In any event, Judge Prado was correct in finding that MALDEF had failed to make a credible claim even if the premise of the disparate-impact approach is accepted.

And, legal theory aside, the approach is bad educational policy. As Abigail Thernstrom wrote in a *New York Times* op-ed (June 10, 1999), "Removing the tests simply shoots the messenger and undermines the drives to raise academic standards." There are racial and ethnic gaps in educational achievement, and those gaps won't be closed by pretending they don't exist or attempting to "litigate them away," as a surprisingly lucid *Washington Post* editorial put it. (December 25, 1999) Instead, competition and accountability among schools should be encouraged through choice, illegitimacy rates lowered (they are around 70 percent for blacks—triple that for non-Hispanic whites), and an end put to the notion that studying hard is "acting white."

Disparate impact theory has always been a bad idea. The focus of a civil rights suit ought to be on whether people of different races are treated differently *because* of their race. That is the commonsense and dictionary meaning of "discrimination," and that is what the 1964 act clearly said and meant. The question of intent, rather than incidental effect, ought to be at the heart of every lawsuit. The ultimate question ought be whether there is actually discrimination—not whether there is failure to achieve racial and ethnic proportionality.

Educators in disparate-impact suits do have the opportunity to rebut the plaintiffs' case by proving that a challenged test is justified by "educational necessity." But it is risky to go to court, trying to prove to a judge or jury—who will know nothing about one's educational enterprise—that the test is a "necessity." Moreover, the technical "validation" frequently insisted on by civil rights plaintiffs, enforcement bureaucrats or federal judges is often impossible. And, conversely, it is almost always possible that a plaintiff in a particular racial or ethnic group can come up with a slightly different test or cut-off score that will diminish the impact on that group while still serving to some extent the educator's end, even if not as well.

In many cases, the use of the disparate-impact approach will result in a federal agency dictating the test. Any educator will want to test students in a way that will not be challenged by the deep-pocketed grantors and litigators from the federal government. Only they can determine what test will meet their approval, and they will be quite happy to share their advice.

But what is really rotten at the core of disparate-impact theory is this: Under the guise of combating the oxymoronic problem of "unintended discrimination," the theory requires deliberate discrimination. It requires tests to be chosen with an eye on the racial and ethnic bottom line. Such a practice would be condemned as discriminatory under any other circumstances—and rightly so.

There are other consequences of the disparate-impact approach that might give its supporters some pause, even if the lowering of standards is unlikely to offend the civil rights establishment.

If it is true, for instance, that Hispanics fail in disproportionate numbers to meet the standards necessary for graduation from high school, then it makes more sense to address this problem directly rather than sweep it under the rug by requiring educators to ignore it. Theoretically, of course, it might be possible to solve the underlying problem while prohibiting tests with a disparate impact, but as a practical matter the latter will undermine the former.

Just about any test is likely to have a disparate impact on some group, whether because of race (remember: whites could sue, too), sex (males can also sue), ethnicity, religion, age or disability—any of which could be asserted as the basis of a federal lawsuit. And that lawsuit is unlikely to be in the interests of every historically aggrieved group.

The use of standardized tests can raise difficult issues, but they are issues for educators and parents, not civil rights lawyers. It is time Congress passed legislation banning the use of disparate-impact theory under Title VI. Schools and parents should be left alone to make educational policy decisions.

References

Chavez, L. Boost of education reform. *Washington Times,* Jan. 13, p. A13.

Clegg, R. The bad law of "disparate impact." *The Public Interest,* Winter 2000, pp. 79-90.

_______. OCR's testing (mis)guidance: anti-education, anti–civil rights. *Civil Rights News,* Winter 2000, p. 1.

G.I. Forum et al. v. Texas Education Agency et al., Civil Action No. SA-97-CA-1278EP, U.S. District Court, Western District of Texas, San Antonio Division.

MALDEF announces it will not be appealing Judge Prado's ruling. (Feb. 8, 2000). *Washington Post* (from news service), p. 9.

Thernstrom, A. Testing, the easy target. *New York Times,* June 10, 1999. p. A31.

Washington Post. When tests have consequences. Dec. 25, 1999. p. A40.

IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF TEXAS
SAN ANTONIO DIVISION

GI FORUM, IMAGE DE TEJAS, §
Plaintiffs 1-7, §
§
Plaintiffs, §
V. §
§ Civil Action No. SA-97-CA-1278EP
TEXAS EDUCATION AGENCY, §
DR. MIKE MOSES, MEMBERS §
OF THE TEXAS STATE BOARD §
OF EDUCATION, in their official §
capacities, §
§
Defendants. §

COMPLAINT

I. INTRODUCTION

1. The Texas Education Agency (TEA) is implementing invalid discriminatory standardized tests as requirements for high school graduation. Under state law, the TEA denies diplomas to Mexican American and African American students at a rate significantly higher than that of Anglo students, without sufficient proof that use of the tests will enhance the education or life opportunities of students. The method of using this test, called the Texas Assessment of Academic Skills (TAAS) exit tests, results in significant and irreparable reduction in the ranks of Mexican American and African American high school graduates. This is occurring and will continue in light of an already high minority drop-out rate. The method of using this test violates a variety of United States Constitutional, statutory and regulatory provisions, as well as fundamental fairness. The implementation of the TAAS exit test in a state with Texas' history of discrimination is particularly counterproductive and violates the orders of the Court in *U.S. v. Texas.*

II. JURISDICTION

2. There is jurisdiction of this case under 28 U.S.C. §1331, 28 U.S.C. §1343, 20 U.S.C. §1706, 42 U.S.C. §2000 (d)(7) and this court's equity jurisdiction to enforce the decrees of the United States District Court for the Eastern District of Texas in *U.S. v. Texas,* 330 F. Supp. 235 (E.D. Tex. 1970), *aff'd,* 447 F. 2d 441 (5th Cir. 1971), *cert. denied,* 404 U.S. 1016 (1974).

III. PLAINTIFFS

3. Plaintiff GI FORUM is an organization dedicated to the educational advancement of Mexican Americans in Texas. They bring this action to ensure that their members' children—Mexican American students in Texas public schools in hundreds of Texas school districts around the state—are not denied an equal educational opportunity to graduate from high school, pursue higher education, join the military or compete in the job market.

4. Plaintiff IMAGE DE TEJAS is an organization dedicated to the educational advancement of Mexican Americans in Texas. They bring this action to ensure that their members' children—Mexican American students in Texas public schools in hundreds of Texas school districts around the state—are not denied an equal educational opportunity to graduate from high school, pursue higher education, join the military or compete in the job market.

5. Plaintiff 1 is a Mexican American student who attended high school in the San Antonio Independent School District. She would have graduated and received a diploma in 1997 but for her failure of the math part of the TAAS test. She has suffered and continues to suffer from the discriminatory policies of the defendants.

6. Plaintiff 2 is a Mexican American student who attended high school in the San Antonio Independent School District. She would have graduated and received a diploma in 1997 but for one point on one part of the TAAS test. Although she had good grades and was on the honor roll for three years, she did not receive a diploma only because of the TAAS. She has suffered and continues to suffer from the discriminatory policies of the defendants.

7. Plaintiff 3 is a Mexican American student who attended high school in the Northside school district for four years. She would have graduated and received a diploma in 1997 but for the TAAS test. She was actively involved in school activities including leadership positions, but failed the math portion of the TAAS. She has suffered and continues to suffer from the discriminatory policies of the defendants.

8. Plaintiff 4 is a Mexican American student who attended high school in an El Paso school district. He would have graduated and received a diploma in 1997 but for one portion of the TAAS test. He has suffered and continues to suffer from the discriminatory policies of the defendants.

9. Plaintiff 5 is a Mexican American student who attended high school in the San Antonio Independent School District. He would have graduated and received a diploma but for the TAAS test. He had good grades and was on the honor roll for two years, but failed the TAAS and did not graduate. He has suffered and continues to suffer from the discriminatory policies of the defendants.

10. Plaintiff 6 is an African American student who attended public

schools in Paris, Texas, who should have graduated in May, 1993. He continued to take the TAAS test at every available opportunity until within the last two years. Because of his age, he is now denied the opportunity to take the test. He completed all requirements to receive a diploma except for the math and reading selections of the TAAS. He has suffered and continues to suffer from the discriminatory policies of the defendants.

11. Plaintiff 7 is a Mexican American student who attended high school in the Harlandale school district for four years. She would have graduated and received a diploma but for one part of the TAAS test. She has suffered and continues to suffer from the discriminatory policies of the defendants.

12. These individual Plaintiffs are representative of the approximately 7,500 students each year who fail the exit level TAAS and do not graduate. These individual Plaintiffs are also representative of the approximately 20,000 to 30,000 members of each sophomore class in Texas schools who drop out before graduation in part because of the TAAS test. These students are denied a diploma, college admission and scholarship opportunities, selection by the military and job opportunities because of the TAAS, regardless of their other qualities, achievements and abilities.

IV. DEFENDANTS

13. Defendants Texas Education Agency, members of the Texas State Board of Education and Mike Moses, as Texas Commissioner of Education have developed and implemented the TAAS, chosen the method of using the TAAS as a graduation requirement, and set the cut-off scores on the TAAS. Individual Defendants are sued in their official capacities. Defendants are the recipients of federal funds.

V. FACTS

A. History of Discrimination Against Mexican Americans and African Americans in Public Schools

14. Mexican Americans and African Americans have suffered from a long and well-documented history of discrimination in Texas public schools. Decades of separate and unequal education have adversely impacted generations of Mexican Americans and African Americans. This past discrimination has consequences in the present, and the Court in *U.S. v. Texas* ordered the state to take affirmative steps to eliminate the vestiges of this past discrimination.

B. What the TAAS Is and How It Is Used

15. First implemented during the 1990-91 school year, the TAAS is now administered in Texas public schools to students in grades 3, 4, 5, 6, 7, 8, and 10. In addition to completing the required high school curriculum, a

student in every public high school in Texas must now pass the reading, writing, and mathematics sections of the exit-level TAAS to receive a diploma. Beginning in the student's spring semester of the tenth grade, the student has eight opportunities to pass the exit level TAAS prior to his or her class scheduled graduation. A student who does not exceed the cut-off score set by the defendants on each of the three parts of the exit level TAAS by the end of his or her senior year is denied a high school diploma even if all other graduation requirements have been met. The student may retake the exam during each subsequent administration of the test, but has no legal right to remedial instruction from any Texas school district if he or she has completed all high school course work. The TAAS is the first state-wide standardized test in Texas to be used to deny high school diplomas to otherwise qualified students.

C. Adverse Effects of the TAAS on Mexican Americans and African Americans

16. The TAAS passage rates of Mexican American and African American first-time takers are significantly lower than that of white students. Whites are almost twice as likely as Mexican Americans and African Americans to pass the TAAS. Although white students have passed the test at a rate of approximately 70 percent, Mexican Americans and African Americans have passed at rates of only around 40 percent. About 60 percent of the minority students in Texas public schools begin their junior years under a cloud of doubt about their futures in the public schools of Texas. They will not be allowed to graduate if they do not pass at least one more part of the test, regardless of their grades and academic record.

17. At the end of every school year, approximately 4,500 Mexican American and 2,000 African American senior students have failed the TAAS and do not graduate. Although Mexican American and African American students make up about 40 percent of Texas high school seniors, they comprise 85 percent of those who fail the last administration of the TAAS.

18. The effects of the TAAS on students of limited English proficiency (LEP) is particularly negative. In the testing of all sophomores in 1995, approximately 11,000 students were identified as Limited English Proficient (LEP). The great majority of these LEP students are Mexican American. Only 14 percent of these LEP students passed the TAAS test the first time they took it. The TAAS exit level test is given only in English even though many LEP students could exceed the performance levels if the test were given in their home language.

19. The diploma denial sanction of the TAAS has had a severely adverse impact on Mexican American and African American students. African American and Mexican American students are far more likely than whites

to be denied a diploma as a result of the TAAS test. This is evident regardless of socioeconomic status, academic track, language program participation and school quality.

20. The TAAS plays a role in the very high dropout rate of minority students in Texas, now approximately 45 percent for Mexican American students and 30 percent for African American students. The Mexican American and African American dropout rates are substantially higher than the drop out rate for white students. The average African American and Mexican American student is significantly more likely to drop out due to the TAAS regardless of socioeconomic status, academic track, language program participation and school quality.

21. In many districts students who fail the test are immediately relegated to academic or educational "tracks" that offer purely remedial education to help them pass the TAAS test, without an opportunity to continue college preparation courses or other appropriate courses for their particular needs. The tracking system related to TAAS is determined primarily at the district level. State regulations require only that students who fail the TAAS be offered some remedial work. Defendants do not prevent districts from requiring students to take only remedial courses or taking so many remedial courses that they cannot timely complete their required course work.

D. Test Validity Issues

22. The TAAS is an invalid instrument for determining which students are qualified to receive diplomas from a Texas public high school.

23. The State of Texas does not provide all students with an equal opportunity to acquire the skills needed to pass the TAAS, including the exit level TAAS. Students do not have equal access to important resources and instruction, and thus a wide gap in preparation opportunity exists between predominantly white school districts or individual schools and predominantly Mexican American or African American districts or individual schools.

24. The TAAS fails properly to assess students' abilities and denies high school diplomas on an inappropriate basis. The test is not appropriately related to what is actually taught or made available to many minority high school students.

25. The inability of the TAAS to properly assess what minority students are actually being taught in high school contributes substantially to both the low minority passing rate and the high minority dropout rate.

26. In Texas, from approximately 1985, state law provided for students to obtain three different types of diplomas—a general, an advanced and an advanced with honors. Separate curriculum and courses were implemented in school districts throughout Texas with separate courses such as "correlated language arts" and "fundamentals of math" replacing courses such as

English and Algebra that were necessary for college and highly related to TAAS exit level passage. The courses in the lower tracks were less likely to contain curriculum necessary for TAAS passage.

27. TAAS tests have been used to place students into remedial classes. As a result of this remedial education, these students have frequently received inferior educations. Low scores on earlier tests have often placed the students who need the most proficient teachers with the least proficient teachers or in less effective curriculum tracks. Instead of improving test scores, "tracking" has contributed to lower and less relevant test scores.

28. The TAAS suffers from technical test design weaknesses that render it unreliable, especially with respect to the writing section of the test. A student who receives a score on the writing assessment cannot reliably be distinguished from a student who receives a score one point higher, yet one point can lead to a denial of a high school diploma.

29. The TAAS contains individual questions that affect different ethnic groups differently, and the inferences made from the test are not justified because they do not sufficiently reflect what minority students are actually learning in the classroom.

30. The language and wording of the TAAS test disfavors LEP students and ultimately reflects such students' abilities to distinguish linguistic subtleties in English rather than their competency on what was actually taught in the classroom.

31. As a predictor of future student performance in the classroom and the workplace, the TAAS is so inaccurate as to render it invalid. There is no proof that TAAS scores differentiate on the basis of characteristics relevant to the opportunities being allocated. There is no or insufficient evidence to show how well TAAS scores reflect real life and educational or job performance. The limited power of TAAS tests to predict success in either school or work means that using test results alone to classify people is discriminatory, especially when test performance is highly correlated with race.

32. The cut-off score used to deny otherwise deserving and qualified students the financial, social, and educational opportunities associated with a high school diploma is arbitrary and capricious. There is no or insufficient empirical evidence to support the contention that students who score at or above the cut-off score on the TAAS are any more qualified or deserving of a high school diploma than those who score below the cut-off score. Many who score below the cut-off score could perform satisfactorily as high school graduates in college, the military and the workforce.

VI. CLAIMS

A. First Claim

33. Defendants have violated their duties under the orders of the United

States District Court for the Eastern District of Texas in *U.S. v. Texas,* 330 F. Supp. 235 (E.D. Tex. 1970), *aff'd,* 447, F. 2d 441 (5th Cir. 1971), *cert. denied,* 404 U.S. 1016 (1974), specifically their duties to ensure that districts are providing equal educational opportunities in all schools.

B. Second Claim

34. Defendants, under color of state law and in violation of the Fourteenth Amendment of the U.S. Constitution, have denied the plaintiffs equal protection of the laws by denying African American and Mexican American students educational and career opportunities equal to those made available to Anglo candidates in violation of 42 U.S.C. § 1983.

C. Third Claim

35. Defendants, under color of state law and in violation of the Fourteenth Amendment of the U.S. Constitution, have denied individual Plaintiffs property and liberty interests in graduating from high school without due process of law in violation of 42 U.S.C. § 1983.

D. Fourth Claim

36. Defendants, recipients of federal funds from the United States Department of Education, have prevented Plaintiffs from graduating from high school and denied them the benefits of a high school diploma. Defendants have subjected the plaintiffs to discrimination on the grounds of race, color, or national origin in violation of Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d et seq.

E. Fifth Claim

37. Defendants, recipients of federal funds from the United States Department of Education, have prevented Plaintiffs from graduating from high school and denied them the benefits of a high school diploma. Defendants have subjected the plaintiffs to discrimination on the grounds of race, color, or national origin in violation of the federal regulations of the U.S. Department of Education implementing Title VI of the Civil Rights Act of 1964, 34 C.F.R. § 100.3.

F. Sixth Claim

38. Defendants, recipients of federal funds from the United States Department of Education, have prevented Plaintiffs from graduating from high school and denied them the benefits of a high school diploma. Defendants have subjected the plaintiffs to discrimination on the grounds of race, color, or national origin in violation of 20 U.S.C. §1703 of the Equal Educational Opportunity Act.

G. Seventh Claim

39. Defendants have denied equal educational opportunity to Plaintiffs on account of their race, color or national origin by failure to take affirmative steps to remove the vestiges of a dual school system, discrimination on the basis of race, color, or national origin in school, and the failure to take appropriate action to overcome language barriers that impede equal participation by its students in its instructional programs in violation of 20 U.S.C. § 1703(f).

VII. PRAYER

WHEREFORE, PREMISES CONSIDERED, Plaintiffs pray that:

1. The Court grant a Declaratory Judgment that the present use of the TAAS exit level test violates the United States Constitutional, statutory and regulatory provisions as alleged in this complaint.
2. The Court enjoin the present use of the TAAS exit level test as a requirement for high school graduation.
3. The Court permanently enjoin Defendants from using any standardized test as an absolute requirement for receipt of a high school diploma.
4. The Court permanently enjoin the defendants' use and method of using the TAAS test until and unless (1) the test is properly validated for the purpose for which it is used and (2) the discriminatory effects of the test, if any, are shown to be as minimal as any reasonably effective alternative, and (3) if the TAAS or any similar standardized test score is used as a factor in determining whether a student may receive a high school diploma, it be used only as any one of several offsetting factors in the determination of whether a student can receive a diploma and that the standardized test score be used as no more than a minor factor in the decision whether to grant a student a diploma.
5. The Court order Defendants to provide compensation to Plaintiffs for reasonable attorneys' fees and costs.
6. The Court grant relief as deemed appropriate by the Court.

DATED: October 14, 1997 Respectfully submitted,

[signature]
ALBERT H. KAUFFMAN
JAVIER N. MALDONADO
NINA PERALES
Mexican American Legal Defense
and Educational Fund, Inc.
140 E. Houston Street, Suite 300
San Antonio, TX 78205
(210) 224-5476
(210) 224-5382 FAX
ATTORNEYS FOR PLAINTIFFS

*Expert Reports**

The Texas Assessment of Academic Skills Exit Level Test

Dr. S.E. Phillips

Michigan State University

A Report Prepared for *GI Forum* *v. TEA*, C. A. No. SA-97-CA-1278EP

January 1999

I. Background Information

The following sections include a career summary, a brief account of prior legal work and a description of my role as a consultant for the Texas Student Assessment Program (TSAP).

A. Career Summary

I have been a member of the graduate faculty in the College of Education at Michigan State University for 16 years and teach courses in educational measurement with a specialization in legal and policy issues. My educational training includes a Ph.D. in educational measurement and statistics from the University of Iowa in 1981 and a law degree in 1990.

My research and scholarship activities have included more than 60 presentations at national professional meetings and 30 papers published in nationally recognized measurement, policy and education law journals. Topics have included standard setting, performance assessment, testing accommodations for persons with disabilities, modifications for English language learners, testing to award diplomas, the Golden Rule remedy, teacher licensure testing and other issues in assessment law.

In 1993, I authored an assessment law handbook for policymakers entitled *Legal Implications of High-Stakes Assessment: What States Should Know.* I have also published eight reviews of standardized assessments and technical measurement texts and regularly contribute a legal issues column for the National Council on Measurement in Education newsletter. A full listing of my presentations and publications is provided in my vita filed in this proceeding.

---

*The expert reports are excerpted with only minor omissions (primarily, references to other witnesses' reports) from the trial testimony and declarations.

I have 20 years of experience working with large-scale assessments in more than a dozen states and several school districts. I have also worked with professional organizations and test publishers on a variety of standardized test instruments. I am currently a member of the Technical Advisory Committees for the Voluntary National Test and for the GED high school equivalency test.

B. Prior Legal Work

I have served as a consultant and expert witness for cases in Alabama, California, Connecticut, Minnesota, Texas and Virginia involving testing accommodations, testing English language learners, test tampering, evaluating teachers, test security, and teacher licensure testing. I have not been deposed for any of these cases and have testified in only two: a due process hearing in Alabama and a district court case in Virginia.[1]

C. TSAP Consultant

I have served as a consultant for the Texas Student Assessment Program since the early 1980s and have worked with the TABS, TEAMS and TAAS assessments. My role as a psychometric consultant has included conducting item response theory workshops for project staff, attending technical advisory committee meetings, reviewing equating results, and providing technical expertise on a variety of assessment issues.

II. Professional & Legal Standards

The major psychometric issues raised by the *GI Forum* lawsuit appear to be primarily related to the use of the TAAS exit level test for awarding high school diplomas. Thus, the information presented in this report focuses on the TAAS exit level test.

In my professional opinion, the Texas Assessment of Academic Skills (TAAS) exit level test meets all relevant professional standards for test development and test use. These standards are enumerated in Chapters 1-5 and Chapter 8 of the 1985 *Standards for Educational and Psychological Testing (Test Standards)* developed and published by three national professional organizations whose members are involved in assessment activities: the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).[2]

It is also my professional opinion that the TAAS exit level tests meet the notice and curricular validity requirements imposed by the Debra P. court.[3] **Adherence to these professional and legal standards has produced a high-quality TAAS exit level test that is valid, reliable and fair for its**

intended use as a graduation test.

In the Preface to the Test *Standards*, the Development Committee stated several guidelines that governed the work of the committee: "The *Standards* should...Be a statement of technical standards for sound professional practice and not a social action prescription.... Make it possible to determine the technical adequacy of a test,...and the reasonableness of inferences based on the test results." (p. v). Recognizing the importance of the *Test Standards*, the Texas State Board of Education specified in its 1995-96 Administrative Code: "The commissioner of education shall ensure that each [test developed according to state statute] meets accepted standards for educational testing." (§ 101.1 (c)).

Under the direction of the Commissioner, the Texas Education Agency (TEA) has obtained input from Texas educators, knowledgeable contractors and national testing experts at important decision points during the development and implementation of the Texas statewide testing program. In particular, for the TAAS exit level test, careful attention has been given to both professional and legal standards for graduation tests. **In my professional judgment, TEA is acutely aware of the high-stakes associated with the TAAS exit level test and has worked diligently with its contractor to develop a quality test that fairly assesses all students.**

Primary & Secondary Standards

The Test Standards are divided into two categories: primary and secondary. "Primary standards are those that should be met by all tests...absent a sound professional reason [to the contrary].... Secondary standards are desirable as goals but are likely to be beyond reasonable expectation in many situations.... Test developers and users are not expected to be able to explain why secondary standards have not been met" (p. 3). The following sections focus on the adherence of the TAAS exit level test to the applicable primary standards for each relevant area.

A. Validity

Validity refers to the weight of accumulated evidence supporting a particular use of test scores. For the TAAS exit level test, scores are used to decide whether students have attained sufficient academic skills in the subject areas of reading, mathematics and writing for the award of a high school diploma. The most important evidence of validity in this situation is a measure of the degree to which the items on each subject matter test measure the knowledge and skills prescribed by the state-mandated curriculum (essential elements). This type of validity evidence is often referred to as *content validity evidence.*

Content Validity Evidence for TAAS. Content validity evidence is typically obtained by professional judgment. Content experts are asked to review each potential test item and classify it according to the objective[4] being measured, check the correctness of the keyed answer, check for ambiguities in wording and other item flaws, evaluate the appropriateness of the content and difficulty for the intended grade level, and identify any inappropriate or potentially offensive language or content. Committees of Texas educators perform these functions for all items written for the TAAS tests.[5] The committees of Texas educators that review the exit level items are chosen to be representative of the state in terms of geography, size of district, gender, and ethnicity. In addition, each committee member is knowledgeable about the grade level and subject matter being tested. Committee members are trained by the contractor and TEA staff prior to beginning their reviews.

As of August 1997, more than 6,000 Texas educators had participated on one or more of the educator review committees for TAAS. During the 1996-97 school year, 16 percent of the item review committee members were African-American and 31 percent were Hispanic.[6] Statewide, the exit level student composition for the Spring 1995 administration was similar (13 percent African-American and 31 percent Hispanic).[7]

All committee-approved TAAS exit level items are field-tested on a representative sample of Texas students prior to use. Field-test items are spiraled within actual test forms to obtain the most accurate data possible.[8] This procedure ensures that student motivation is as high for field-test items as for those that count in their scores. From the field-test data, various statistics are calculated to summarize student performance on each field-test item. Included at this stage are measures of differences in performance between majority and minority group students.

Prior to the construction of final test forms, all field-tested items are reviewed again by the educator committees with particular attention to those items identified as having large differences between the performance of African Americans and whites or Hispanics and whites. Items with context or language characteristics that the committee believes may be contributing to the differential performance are revised and field-tested again or are dropped from further consideration.

In addition to convening educator committees to evaluate the content validity of each potential TAAS item, TEA staff also conduct reviews to ensure that each test form is representative of the state objectives it measures. For each subject area, TEA, with input from the Texas educator committees, has prepared a test blueprint which describes the mix of content and skills to be tested by each exit level form. As each new exit level form is constructed, items are chosen to match the specifications contained in the test blueprint. In addition, based on field-test information, each new form is constructed to have statistical properties that are very similar to prior forms. This process is referred to as "constructing parallel forms" and theoretically should result in forms that are so similar in their difficulty and content that a student given a choice would be indifferent about which form to take.

Other Validity Evidence. Other types of validity evidence include criterion and construct validity evidence. Criterion validity evidence, in the form of correlation coefficients, is most appropriate for situations in which test scores are used to predict outcomes such as freshman grade point averages. It can also be useful in determining the degree to which two tests measure the same or different skills. Because TAAS exit tests are intended to measure state-specific content knowledge and skills, and not to predict any other outcome, criterion validity evidence is tangential.

Construct validity evidence refers to the sum of research knowledge and experiments designed to define a psychological construct, such as extroversion or locus of control, that an instrument is intended to measure. Because the TAAS exit level tests are designed to measure specific academic content, not to define more general psychological constructs, construct validity evidence is also tangential in this context.

B. Reliability

Reliability is an indicator of consistency of measurement. Errors of measurement are minimized and decision consistency is maximized by a reliable test. Reliability is a necessary but not sufficient condition for validity.

There are two major procedures for calculating test reliability: repeat testing and measures based on a single test administration. **Repeat testing is impractical for the TAAS exit level test for two reasons: (1) decreased student motivation on a second testing that doesn't count alters performance; and (2) schools are unwilling to devote additional instructional time to unnecessary double testing of students.** Thus, TEA reports reliability measures based on a single test administration. These measures are called KR_{20} reliabilities and are reported as decimal values between zero and one. *A common rule of thumb for a test used to make decisions about individual students is to require a reliability of at least 0.85.*

One way to compute reliability for alternate forms of a single-administration test is to split the test into two parallel halves. The KR_{20} reliability estimate is an average of all such possible splits so it includes errors related to item sampling. Sources of error due to testing at different points in time are included in retest reliabilities but not KR_{20} reliabilities. However, because students are expected to continue receiving instruction between test administrations, one would not expect TAAS exit level test scores to remain constant over time. **Thus, KR_{20} values are the most appropriate reliability estimates for the TAAS exit level tests**.

TAAS Reliabilities by Ethnicity. TAAS exit level reliabilities are based on the entire population of students tested. For the 1997 spring administration of the exit level test, over 200,000 tenth graders (about 27,000 African-Americans and 68,000 Hispanics) took the exit level test for the first time. Reliabilities by ethnic group are presented in Table 1.[9]

The data indicate that all reliabilities for the reading and mathematics tests on which passing decisions are made are high and exceed the .85 rule of thumb. **Reliabilities for African-Americans and Hispanics are higher than for whites for all subject areas.**

Table 1
TAAS Exit Level Reliabilities

Ethnic Group	Reading (48 items)	Mathematics (60 items)	Writing (40 MC items)
African-American	.88	.94	.83
Hispanic	.89	.94	.86
White	.86	.92	.81

The reliabilities for writing presented in Table 1 are for the multiple-choice portion of the test. Writing multiple-choice scores are combined with an essay score to produce a writing total score. Essays are scored holistically on a four-point scale and scorer agreement after three readings is 98 percent.[10]

Standard Error of Measurement at the Passing Score. The standard error of measurement (SEM) is derived from the test reliability and has the same metric as the test score. In the Reliability Chapter of the *Test Standards,* a secondary standard recommends reporting the standard error of measurement at the passing score. For the TAAS exit level tests, the standard errors of measurement at the passing scores are approximately 2 -3 raw score points, about the same values as the overall standard errors of measurement reported in the *Technical Digest.*[11]

Errors Due to Multiple Retakes. Measurement errors are assumed to be random. Sometimes such errors will be positive and benefit the student, while at other times measurement error will be negative and disadvantage the student.

These two types of measurement error are referred to as false positives and false negatives. A **false positive** occurs when *positive measurement error* results in a passing score for a student whose true achievement is below the passing standard. Such students pass the test even though they have not actually attained the required level of achievement. A **false negative** occurs when *negative measurement error* results in a *failing score* for a student whose *true achievement is above* the passing standard. Such students must retake and pass a different form of the test to earn a high school diploma.

For a student's first attempt to pass the TAAS exit level test, the probability of *a false positive result is modest. However, for students who take advantage of the full eight attempts to pass the TAAS exit level test available prior to their scheduled graduation, the probability of a false positive error beneficial to the student is substantial.*

For example, the 1997 TAAS exit level Reading Test consisted of 48 items, had an approximate standard error of measurement of 2 raw score points, and a passing score of 34. A student with a true achievement (no measurement error) of 33 (about one-half standard error below the passing standard) had about a one in three chance of passing the test on the first attempt. After eight attempts, the student's chances of success rose substantially to more than nine chances out of 10. **This means that a student whose true achievement was one-half standard error below the passing standard in 1997 had extremely high odds of passing the reading test after multiple retakes without receiving any remediation.**

A student with true achievement one standard error below the passing score had substantial 75 percent chance of passing the TAAS exit level reading test after eight attempts with no intervening remediation. That is, out of 100 students with true ability two points below the passing score, approximately 75 would pass the test after eight attempts due to help from random positive errors of measurement. If these students also received intensive remediation as required by state law, their true achievement would increase and the probability of passing on a subsequent attempt would increase even more dramatically.

Conversely, for students with true achievement at the passing score, the probability of passing after eight attempts is near certain. For students with true achievement one standard error above the passing score, the probability of passing after eight attempts is virtually 100 percent.

Relating the Passing Score to the SEM. Some professionals have advocated an alternative passing standard that is three standard errors below the passing score set by a policy-making board. The rationale for this recommendation is to minimize false negatives. This argument might have some merit if passing decisions were being made based on a single attempt because negative errors of measurement could cause a student with true ability at or slightly above the passing score to fail a single administration of the TAAS exit level test. However, students in Texas have eight attempts to pass the TAAS exit tests prior to graduation.[12] These multiple attempts make a false negative an extremely rare event.

After eight attempts, virtually all students with passing scores at or above the passing standard will achieve a passing score and a substantial proportion with true achievement one to three standard errors below the passing score will also pass. For example, if there were 100 students each with true ability 1, 2 and 3 standard errors below the passing score, after eight attempts, (88+55+27) / 300 = 170/300 = 57 percent would pass the test.

These would all be erroneous passing decisions for students whose true achievement had not changed while there would be virtually no erroneous failures for students who actually had attained the required level of achievement.

In sum, with regard to errors of measurement on the TAAS exit level tests, the availability of multiple retakes provides a substantial probability of errors in the students' favor (false positives) and a negligible chance of errors disadvantageous to students (false negatives). Therefore, lowering the passing score to prevent a minute number of potential false negatives is not justified when compared to the large number of additional false positives that would be created. **To the extent that more minority than majority students attain TAAS exit level scores 1-2 SEMs below the passing score, more minority students are likely to benefit from positive errors of measurement.** While false negatives are corrected via repeat testing, false positives are neither identified nor corrected. That is, *a student who fails erroneously is given another chance to pass while a student who passes erroneously is allowed to retain the benefits of an unearned passing decision.*

C. Test Development and Publication

The Test Development and Publication chapters of the *Test Standards* charge test developers with the responsibility for following professionally accepted procedures for test construction and for disseminating information that promotes appropriate test use.[13] The procedures for designing the TAAS exit level test to measure the state objectives are summarized in the Validity section of this report and in greater detail in the *1996-97 Technical Digest.* The consensus process used to specify the content to be included in the state objectives and a listing of the individual objectives and instructional targets for the reading, mathematics and writing TAAS exit level subtests are also included.[14] **The TAAS test construction process is detailed, comprehensive, sensitive to concerns from diverse groups, and consistent with industry standards.** All scored items on the TAAS exit level test are released to the public annually.

Multiple methods are used to encourage appropriate TAAS test preparation and use of results. Educators who participate on review committees and school personnel who administer the TAAS tests are required to sign confidentiality and security maintenance agreements. The *Texas Administrative Code* describes and lists conduct that is prohibited because it would compromise the integrity, validity and fairness of the TAAS tests.[15] The confidentiality of individual student data is also protected.[16] Appropriate score uses and cautions for score use are included in the *Technical Digest.*[17] *Score reports and their accompanying interpretive materials have been designed to facilitate appropriate interpretations and uses of TAAS data.* TEA staff regularly respond to questions from Texas educators.

D. Technical Characteristics

Current technology implemented by experienced personnel is utilized to maintain the technical quality of the TAAS exit level test. Score comparability across test administrations is high ensuring that the graduation standard remains constant for all students.

Equating and Scaling. Test items measuring the same content will differ in difficulty; some test items on a given topic are easier for students to answer correctly and others are harder. Well-constructed parallel forms of the same test instrument will have approximately the same level of difficulty. To adjust for any remaining minor variations in difficulty, test forms are equated to a common scale. This ensures that the passing standard, which is specified on the common scale, will be the same for all students no matter when they were tested or which form they were administered.

The Rasch Model. The Rasch item response theory model is used to equate forms of the TAAS exit level test. **The Rasch model is a professionally recognized method that has been used on achievement tests for over 25 years.** This model has been used successfully by several states and a national test publisher to equate and scale large-scale, standardized achievement tests.

The Rasch model is especially well-suited to statewide testing because it allows properties of items to be compared on the same scale regardless of which subgroup of students responded to the item. It is also a parsimonious model because it captures the primary information available for a test item in a single parameter, the item difficulty. Item difficulties range from about -3 to +3 (like z-scores). Hard items have positive values; easy items have negative values.

The Rasch model makes the following assumptions: unidimensionality, local independence, equal item discrimination and zero guessing. *These Rasch model assumptions are appropriate and reasonable for the TAAS exit level test.*

Unidimensionality means that the test must be designed to measure a single trait. Reading, mathematics and writing have been shown to be adequately unidimensional traits for obtaining good results with the Rasch model. *Local independence* means that the answer to one item on the test does not depend on answers to other items on the test. Although small groups of TAAS items may relate to the same passage or graphic, they are carefully constructed to measure independent skills. That is, an answer obtained for one item is not used or related to the correct answer for any other items in the set. Therefore, the TAAS test satisfies the local independence assumption.

Item discrimination is a measure of the degree to which high-scoring students tend to answer the item correctly and low-scoring students tend to answer the item incorrectly. Item discrimination values vary across items, but for well-constructed tests the variation is contained within a relatively small interval. For large student populations like Texas and well-constructed instruments like the TAAS exit level test, the Rasch model is robust to the relatively small variations in item discrimination that occur. This means that the Rasch model produces accurate equating results in spite of variations in item discrimination.

The *zero guessing* assumption means that the model assumes that students do not obtain correct answers to items by random guessing. This is a reasonable assumption for the TAAS exit level test for two reasons: (1) students typically have at least some partial knowledge on which to eliminate one or more answer choices from consideration (i.e., they are not guessing randomly); and (2) if random guessing were occurring, it should be distributed evenly across answer choices. Item data for achievement tests indicate that the percent of students who could have obtained a correct answer by random guessing is extremely small because the incorrect answer choice chosen least often by students usually has a response percentage that is less than 10 percent.

Choosing the Rasch Model for TAAS. The Rasch model was chosen from a family of item response theory (IRT) models that can be used to equate and scale large-scale achievement tests. In addition to item difficulty, the more complex IRT models also estimate discrimination and guessing parameters for each item. But these additional item parameters are not always estimated accurately, even when the number of students tested is large.

Use of the Rasch model provides equating results that focus on the item difficulty parameters, which contain the majority of important item information and are most accurately estimated. When other parameters are added to the model to account for item discrimination and guessing, they add a lot of noise to the system because they often contain relatively large estimation errors or are assigned a default value due to too little data being available for estimating their values.

Moreover, *the more complex models base their measures of student achievement on differential item weighting.* This means that two students who achieve the same raw score will receive different scaled scores if they correctly answered different subsets of test items. However, **in the Rasch model used to equate the TAAS exit level tests, student performance results and passing decisions are based on the student's raw score (number of items answered correctly).**

Equating and Pre-equating. The passing standard for the TAAS exit level test was set on a base form given the first year the test became operational. The Rasch model has been used to equate all subsequent forms to the common scale of the base form. If a new TAAS form is more difficult than the base form, fewer correct answers are required to pass. If a new

TAAS form is easier than the base form, more items must be answered correctly to pass the test. However, **intensive test development efforts have produced TAAS exit forms that are extremely similar in content, difficulty, and reliability. Thus, equating constants have been extremely small resulting in only very minor adjustments to the original raw score passing standards.** In my professional experience, errors associated with Rasch model equating are generally within ½ raw score point, which is small relative to the overall standard error of measurement of about 2-3 raw score points.

In addition to maintaining equivalent passing standards via equating, the Rasch model is also used to pre-equate TAAS test forms. This is a more accurate procedure than relying on sample statistics because Rasch item difficulties are not dependent on the characteristics of the particular sample of students who responded to the item. Pre-equating uses field-test data to make test forms more similar by keeping the average difficulty of the items for each objective as comparable as possible and by minimizing the potential differences in raw score passing standards between forms.

In sum, the simplifying assumptions used in the Rasch model are justifiable for achievement tests such as the TAAS exit level test and provide a powerful tool for ensuring fairness for all students.

Passing Standards. The responsibility for setting passing standards on the TAAS exit level test resides with the State Board of Education. The Texas Education Code states: "The State Board of Education shall determine the level of performance considered to be satisfactory on the assessment instruments."[18]

The *Test Standards* require that the procedures used to establish the passing standard on a graduation test be documented and explained but do not require any specific method to be used. Documentation provided by the contractor and contained in the *Technical Digest* indicates that educator committees provided recommendations to TEA and the commissioner. The commissioner in turn provided a recommendation to the State Board that included field test estimates of passing rates at passing standards of 60 percent and 70 percent correct. The State Board made the final decision to set the passing standard at 60 percent for the first year and at 70 percent thereafter.[19]

With minor modifications, the TAAS exit level test was constructed to measure the same state level essential elements as the TEAMS graduation test that preceded it. The major difference between the TEAMS and TAAS graduation tests is the level and complexity of the skills assessed. The TEAMS test focused on basic skills; the TAAS test covers the same curricular areas but measures them at a higher level and places more emphasis on higher-order thinking and problem-solving skills. Thus, by design, the TAAS exit level test is more difficult than the TEAMS test.

For their discussions with the Commissioner regarding passing standards for the new TAAS exit level test, TEA received input from the educator committees that reviewed the specifications and items for the more difficult TAAS test.[20] They also had results from an equating study which related TAAS scores to their equivalents on the TEAMS scale. This information, together with student performance data from the field test, provided the basis for the Commissioner's recommendation to the State Board.

As was pointed out to the State Board members, field-test estimated passing rates must be viewed cautiously because they represent student performance under conditions of low motivation. As expected, student performance on the TAAS exit level test increased significantly from field-testing to the first live administration. Thus, in weighing the goal of increasing the academic proficiency of high school graduates in Texas and data known to be an underestimate of student performance to be expected when TAAS was fully implemented, it was reasonable for the State Board to choose to phase-in the 70 percent passing standard.

Nothing in the law, administrative code, or *Test Standards* prescribes what information State Boards should consider or how they should weight the information in arriving at a passing standard. **The State Board acted lawfully and within its authority when it established the 70 percent passing standard.** TAAS exit level data clearly indicate that substantial numbers of students in all ethnic groups are meeting this standard on their first attempt and that remediation for nonpassing students has been successful.

In addition, the *Texas Administrative Code* provides: "On the [exit level test], a student shall not be required to demonstrate performance at a standard higher than the one in effect when he or she was first eligible to take the test."[21] To satisfy this mandate, TEA still administers the TEAMS test to those individuals who left school without a high school diploma during the years that the TEAMS test was required, even though nearly a decade has passed since the TEAMS test was replaced by TAAS.[22]

E. Legal Requirements

In the *Debra P. v. Turlington* case, the court instituted two additional requirements for graduation tests: *notice* and *curricular validity*. The curricular validity requirement, also referred to as *opportunity to learn,* was included in the 1985 revision of the *Test Standards.*[23]

Notice. Notice requires the state to disseminate information about graduation test requirements to all affected students well in advance of implementation. This responsibility is codified in the *Texas Administrative Code* as follows:

> The superintendent of each school district shall be responsible for the following: (1) notifying each student and his or her parent or guardian in writing no later than the beginning of the student's seventh grade year of the

> essential skills and knowledge to be measured on the exit level...tests administered under the [Texas Education Code]; (2) notifying each 7th-12th grade student new to the district of the testing requirements for graduation, including the essential skills and knowledge to be measured; and (3) notifying each student required to take the exit level . . . tests and out-of-school individuals of the dates, times, and locations of testing." [§ 101.2(a)].

The notification provided to students and their parents occurs more than three years before the first TAAS exit level tests are administered in the spring of 10th grade and more than five years prior to the expected graduation of these students in the spring of the 12th grade.

Opportunity to Learn. Opportunity to learn (OTL) means that students must be taught the skills tested on a graduation test. In practice, evidence of OTL is gathered by examining the official curricular materials used in instruction and by surveying teachers to determine whether they are teaching the tested content. **For the TAAS exit level tests, OTL has been established through the state-mandated essential elements and adequacy of preparation reviews by Texas educator committees and separate bias review panels.**

In the *Debra P.* case, the court held that the appropriate standard for instructional validity is that "the [tested] skills be included in the official curriculum and that the majority of the teachers recognize them as being something they should teach."[24] The *Debra P.* court also found that:

- even if the present disproportionate failure rates [on the Florida graduation test] were caused by past discrimination, the state had adequately demonstrated that [the graduation test] was a necessary remedy;
- it was not constitutionally unfair that some students had mediocre teachers; and
- proving instructional validity for each individual student was an impossible burden.[25]

State Mandated Content. The Texas Education Code provides: "The State Board of Education by rule shall establish the essential skills and knowledge that all students should learn..." (§ 39.021). Representative committees of Texas educators, business representatives, parents and the public participated in the establishment of the state essential elements tested by the TAAS exit level test. By law, all Texas public schools are required to teach this content and to provide remediation to unsuccessful students. The essential elements and state objectives have been widely disseminated to Texas educators, students, parents and the public.

Adequacy of Preparation Reviews. As indicated earlier, all TAAS exit level test items are reviewed by committees of Texas educators representative of the ethnic composition of exit level students. As part of these item reviews, each participating teacher is specifically asked to judge the adequacy of preparation of exit level students for demonstrating the academic skills required to correctly answer each item.[26]

In addition to being asked to judge whether each TAAS exit level item is a good measure of the curriculum and suitable for 10th-grade students, the teachers on the item review committee are also asked to respond "yes or no" to the following question for each test item: "Would you expect students in your class to have received sufficient instruction by the time of the test administration to enable them to answer this item correctly?" Large majorities of committee members respond "yes" for all TAAS test items included on exit level forms.

In the early years of the TAAS exit level test, additional adequacy of preparation judgments were obtained from separate bias review panels composed entirely of minority educators. These minority educators specifically considered whether minority exit level students had an adequate opportunity to learn the tested content, and they were supportive of the TAAS exit level test.

Furthermore, because the TAAS exit level test was constructed to measure the same essential elements as the TEAMS test which preceded it, the adequacy of preparation surveys of Texas educators conducted on the essential elements and test items for TEAMS were also useful in documenting opportunity to learn for the TAAS exit level test.

Remediation. The *Debra P.* appeals court stated: "[The state's] remedial efforts are extensive.... Students have five chances to pass the [graduation test] between 10th and 12th grades, and if they fail, they are offered remedial help.... All [of the state's experts] agreed that the [state's remediation] efforts were substantial and bolstered a finding of [adequate opportunity to learn]."[27]

The *Texas Education Code* provides: "Each school district shall offer an intensive program of instruction for students who did not [pass the TAAS exit level test]." [§ 39.024 (b)].

Study Guides. The *Texas Education Code* provides: "The agency shall develop [and districts shall distribute] study guides for [the TAAS exit level tests] to assist parents in providing [summer help to students who fail a TAAS exit level test]." [§ 39.024 (c)]. These guides have been developed and distributed.

Released Tests. When the TAAS exit level tests were initially implemented, TEA provided all Texas school districts with sample test items for each subject area. In 1995, the *Texas Education Code* was amended to provide for the annual release of all scored TAAS exit level test items.[28] Teachers and students can use this information to review and practice for subsequent test administrations.

Collectively, the (1) well-publicized, state-mandated TAAS exit level objectives that all schools are required to teach; (2) wide dissemination to

students, parents and educators; (3) positive adequacy of preparation reviews by educator committees and bias review panels; (4) mandated remediation; (5) distribution of study guides; and (6) availability of released tests, provide strong evidence of adequate notice and opportunity to learn for the TAAS exit level tests. This evidence also demonstrates that Texas has instituted a comprehensive support system for all students subject to the TAAS exit level graduation test requirement.

III. Differential Performance

Differential performance occurs when passing rates for African-American and Hispanic students (minority groups) are lower than the passing rates for white students (majority group). When the differential performance between minority and majority groups becomes too great, it is labeled adverse impact. An important issue in this context is determining when differential performance becomes large enough to qualify as adverse impact.

In employment testing, two types of significant differences are commonly used to assess adverse impact: practical significance and statistical significance. Statistical significance is important when the group differences being used to evaluate potential adverse impact represent samples from their respective populations. In such cases, the relevant question is whether the sample differences are the result of random error or true population differences. Statistical tests can be used to evaluate whether the differential performance among the samples is large enough to justify the conclusion that there is differential performance among the respective minority and majority populations.

Once differential performance has been established for a minority population, one must decide if it is large enough to justify labeling it adverse impact. This requires a judgmental evaluation of the practical significance of the population differences. The *Uniform Guidelines* for employment testing label differential performance as adverse impact when the passing rate for the minority group is less than 80 percent of the passing rate for the majority group.[29]

For large-scale, statewide graduation tests such as the TAAS exit level tests, statistical tests for evaluating adverse impact are unnecessary because the reported passing rates are based on the entire population of students tested in each ethnic group. Using statistical tests designed for samples is inappropriate when population values are known.

TAAS Passing Rates. No statistical tests are needed to determine that the white initial passing rates exceed those for African-American and Hispanic students for all years and subjects. There are three important questions to be considered in evaluating the differential performance among these populations:

1. Is the differential performance between the minority and majority populations of sufficient practical significance to warrant the label "adverse impact"?
2. Is a different conclusion warranted when cumulative passing rates are compared?
3. Do the trends in minority student performance indicate that the education of minority students has improved in Texas?

The TAAS exit level test data support a "yes" answer to each of the three questions. The specifics are presented in the next three sections.

1. Practical Significance of Differential Initial Passing Rates

The overall initial passing rates for the first attempt in 10th grade for African-Americans and Hispanics are below the 80 percent white passing rates for all years. This suggests that the differential performance between the minority and majority groups is of sufficient magnitude to be labeled adverse impact.

The passing rates for all three groups increased over the period 1994 to 1998 and that the largest gains were made by African-American and Hispanic students. The percent increase in passing rates was greatest in mathematics where African-American and Hispanic passing rates increased 85 percent and 63 percent, respectively, compared to only a 26 percent increase for whites over the five-year period.

From 1994 to 1998, both minority groups also closed the gap between their passing rates and the 80 percent standard. African-Americans moved from 25 points below the 80 percent standard in 1994 to 13 points below in 1998. The Hispanic group closed the gap from 19 points below the 80 percent standard in 1994 to 9 points below the standard in 1998. Overall, the African-American initial passing rate rose 26 points during this four-year period while the Hispanic initial passing rate rose a total of 24 points.

2. Cumulative Differential Passing Rates

As indicated previously, according to state law, Texas students who do not pass the TAAS exit level test on the first attempt are entitled to intensive remediation provided by the district. These students have a total of eight attempts to pass the TAAS exit level test prior to their scheduled graduation. **Therefore, with respect to adverse impact, the focus should be on the cumulative passing rates across all attempts prior to graduation.**

The overall cumulative TAAS passing rates for African-Americans and Hispanics exceeded the 80 percent standard for the Class of 1996, the Class of 1997 and the Class of 1998. Although the initial passing rates for minority students met the 80 percent standard for adverse impact, the cumulative TAAS passing rates for those same minority groups did not.

Over time, despite their initial disadvantage in skill level, significant numbers of minority students have overcome their academic weaknesses and succeeded on TAAS.

3. Educational Improvement of Minority Students: Benefits for the Class of 1997

For the Texas 12th-grade students in 1997, 790 out of every 1,000 Hispanic students had passed the TAAS tests required for graduation. The initial passing rate for these Hispanic students when they were 10th-graders in 1995 was 370 out of every 1,000 students.

The combined dropout rate for Hispanic students in 1996 and 1997, the two years between their initial TAAS attempt in 1995 and their expected graduation in the spring of 1997, was 52 out of every 1,000 students. Some of these Hispanic students may have dropped out due to academic difficulties while others dropped out due to nonacademic reasons (e.g., family illness, employment, military). We do not know what percent of the Hispanic students who chose to drop out had not yet passed TAAS.

For purposes of illustration, assume that 50 percent of the dropouts had not yet passed the TAAS exit level test. Then approximately (370-26)/(1000-52)*1000 = 363 per 1,000 remaining Hispanic students passed TAAS on the first attempt. Since 790 out of every 1,000 Hispanic students had passed TAAS after eight administrations, approximately 790 - 363 = 427 out of 1,000 received sufficient remediation to pass TAAS on a subsequent attempt.

Given the 1997 12th-grade enrollment of (195,075)(.374) = 72,958 Hispanic students, approximately 31,153 Hispanic students statewide who had not attained the state objectives in 10th grade had received sufficient remediation to do so by the time of their expected graduation in the spring of 1997. Similar calculations for African-American students yield an estimate of 13,362 remediated students. **Altogether, about 44,515 minority students in the Class of 1997 were successfully remediated after having failed their first attempt to pass the TAAS exit level test in the spring of 1995.** Had these 44,515 minority students not taken TAAS in 10th grade, it is unlikely that their skill deficiencies would have been identified and remediated.[30]

Cost/Benefit Analysis. In this scenario, the ratio of students remediated to nonpassing dropouts is 16:1 for Hispanic students and 21:1 for African-American students. That is, for every Hispanic student who may have dropped out of school due to academic problems identified by the TAAS exit level test, 16 were successfully remediated; for every African-American student who did so, 21 were successfully remediated. In a cost/benefit sense, the number of minority students benefiting from the TAAS exit level test clearly outweighs the few who may have given up in discouragement after a poor performance on their initial TAAS attempt.

It should also be noted that schools have difficulty remediating students who choose to drop out and that dropping out is a legal option for students after age 16 (approximately 10th grade). While eliminating the TAAS graduation requirement might induce some students to remain in school, it would decrease the value of the high school diploma as an indicator of skill attainment, especially for minority students. Moreover, for those students who drop out of school due to nonacademic reasons, elimination of the TAAS graduation test would have no effect. A more efficient and direct solution for keeping all students in school through the 12th grade would be a statutory change increasing the legal age for leaving school to age 18.

Fairness. For the Class of 1998, the cumulative passing rates for African-Americans and Hispanics were 82 percent and 83 percent, respectively. Based on statewide information for all grades tested, an additional 6 percent to 10 percent of these minority students may have been exempted from passing TAAS based on their individualized educational plans. That leaves only about 8 percent to 11 percent of the minority students unaccounted for. Some may not have completed all courses required for graduation in their districts, and some may have passed TAAS at a subsequent summer or fall administration.

When the TAAS exit level tests identify students who have not attained the state mandated objectives and schools successfully remediate those students, the result is high school graduates with higher skill levels than they would have attained had their deficiencies not been identified. *Would it be fair to the 82 percent to 83 percent of African-American and Hispanic students from the Class of 1998, who worked hard to attain the skills needed to pass the TAAS exit level test, to allow the 8 percent to 11 percent of minority students who were not successful on TAAS to also receive a high school diploma?* A judge in the *Debra P.* case put it this way: "It is undoubtedly true that the appearance of having been educated may be accomplished by the conferring of a diploma. Nevertheless, if [the student has not learned the tested skills], even the most emphatic judgment and order of the most diligent court cannot supply [the missing achievement]."[31]

If those minority students who were unable to pass the TAAS exit level test were awarded a high school diploma by court order, these students would be erroneously certified as having satisfactory educational attainment. It is likely that the benefits for these students would only be temporary; an employer relying on the diploma would certainly discover the lack of skills during the probationary period and discontinue the employment.

In the interim, these students would have been given false hopes of a better job and would face a losing battle to retain jobs for which they were not fully qualified. Further, for those minority students with high school diplomas who were qualified, employers might use their experiences with unskilled diploma holders to discount the credentials of all minority applicants in a return to the stigmatizing assumption that minority students are incapable of achieving at the same level as white students. Consequently, minority students who passed the TAAS exit level test and those who did not would both be hurt by a court-ordered reversion back to a system of awarding high school diplomas based on seat time and social promotion.

Multiple Measures. It is important to note that passing the TAAS exit level test is not the only requirement for receiving a high school diploma in Texas. Students must also pass all of their required courses and meet any additional requirements imposed by their school districts. Students are required to meet both testing and course requirements because each represents a different kind of accomplishment that is valued in a high school graduate.

Moreover, students who fail a single course may be unable to graduate on time just as those who do not pass the TAAS exit level test may have to delay graduation. And in both cases, students have multiple opportunities to complete the failed course or retake the failed TAAS subtest. Furthermore, a student who is not awarded a high school diploma due to not having passed one or more TAAS subtests has not been denied a diploma based on a single piece of data. Rather, the denial is based on at least eight scores from eight forms of TAAS administered on eight different occasions.

Compensatory Measures. There are some advocates who argue that course grades should be considered for those students who are unable to pass TAAS after several attempts. Doing so would create a compensatory model in which passing grades in courses with low level or unrelated content could offset a student's failure to achieve the state objectives.

Alternatively, the grade the student earned in a particular content course might have been based in part on factors other than achievement (e.g., attitude, effort, improvement). If so, it would not be appropriate to allow success on those factors to compensate for lack of achievement of the state objectives. In sum, grades are not equivalent measures of the state objectives measured by the TAAS exit level test and may reflect lower standards and rewards for seat time. Therefore, grades should not be allowed to compensate for a student's inability to pass the TAAS exit level test.

Other Indicators of Improving Minority Achievement. In addition to improved passing rates on the TAAS exit level tests, there are several other indicators of improved educational attainment for African-American and Hispanic students in Texas, including substantial improvement in the percent mastering all TAAS exit level objectives and in average SAT scores for African-American and Hispanic students.[32]

Some schools have shown significant increases in the percent of students passing all TAAS tests in three Texas secondary schools with substantial minority and economically disadvantaged students. These schools were identified in a research study conducted by the Dana Center at the University of Texas at Austin which commended them for high levels of academic success in poor communities.[33] Schools selected for the study had at least 60 percent economically disadvantaged students and at least 70 percent of students passing the TAAS reading and mathematics subtests.

A followup study of 11 elementary schools, located primarily in the Rio Grande Valley, is being conducted to determine whether their academic success with high poverty Limited-English proficient (LEP) students can be replicated in other schools. Schools selected for this study have the following characteristics: at least 40 percent LEP students, at least 50 percent economically disadvantaged students, no LEP TAAS exemptions, and a recognized or exemplary rating.[34]

Differential Item Performance. When minority and majority students exhibit differential levels of performance on an achievement test, some observers want to believe that the test items are "biased" against members of the lower-scoring minority group. However, an equally plausible explanation for the differential performance is a true difference in average achievement levels for the two groups.

To investigate the possibility that differential item performance is the result of item characteristics that unfairly disadvantage a specific minority group, two analyses are completed for each TAAS test item. **First, a statistic is calculated which quantifies differential item performance for minority and majority groups of *equal ability*.** Basing these item comparisons on minority and majority groups of equal ability eliminates the possibility that any observed differences are due to achievement differences between the two groups.

Second, the differential item performance statistics are reviewed by panels of content experts with proportional minority membership. Particular attention is given to the items with the largest differential performance statistics because they are least likely to have been caused by random errors in the statistical procedure. Great deference is given to the views of committee members from the minority group exhibiting the differential performance. An item that exhibits statistically significant differential performance between minority and majority students can be retained for use on a TAAS test only if, in the professional judgment of the item review committee, the item is a fair measure of its corresponding state objective *for all students,* and is free of offensive language or concepts that may differentially disadvantage minority students.

IV. Dropout Data

Society benefits when students stay in school and earn a high school diploma because high school dropouts typically hold lower paying jobs and have limited opportunities for advancement.[35] Nationally, there is a growing concern about differential dropout rates among ethnic groups, particularly for Hispanic students who have the highest rates for leaving school without a high school diploma.[36]

Texas Dropout Data by Ethnicity. In Texas, consistent with national trends, dropout rates for Hispanics and African-Americans exceed those for whites. Dropout data for 1996-97 by grade and ethnicity indicate that annual dropout rates for African-Americans are about 1 percentage point higher than for whites; for Hispanics, the annual dropout rate is about 2 percentage points higher than for whites. Nonetheless, these data also indicate that the vast majority of students in all groups are staying in school; Texas longitudinal dropout estimates for grades 7-12 suggest that about 88 percent of minority students and about 94 percent of majority students remained in school in 1996-97.

Moreover, for each ethnic group, the annual dropout rate in Texas is significantly less than the corresponding rate nationally. In a 1994-95 government study of high school dropout rates in 29 states, including California, New York, and the District of Columbia, Texas ranked second lowest behind North Dakota.[37]

Some advocates in Texas have blamed the TAAS exit level test for the larger minority dropout rates. However, the data do not support this assertion.

Students first attempt the TAAS exit level test in 10th grade. If anticipated or actual failure on this test caused substantial numbers of minorities to drop out of school, one would expect a spike in the number of dropouts in 10th and 11th grades. The data indicate no such spike. Dropout rates for all groups are relatively flat in 10th and 11th grades. **The largest percentage of dropouts occurs in 12th grade for African-Americans and in ninth grade for Hispanics, well after and well before the first TAAS attempt.**

Historical Dropout Trends. Historical trends in annual and longitudinal dropout rates also do not support the assertion that TAAS implementation caused dropout rates for minority groups to increase. **Since the implementation of the TAAS exit level test in 1990, dropout rates for African-American and Hispanic students have steadily declined, and the gap between minority and majority students has shrunk from about 15 points longitudinally in 1990 to about 6 points in 1997.** There is no evidence that introduction of the TAAS exit level test affected the dropout rate for any group.

Dropout Characteristics. The percents of total enrollment and percents of dropouts, compared for different populations of Texas students, indicate that economically disadvantaged, at-risk, Title I, special education, and bilingual students all drop out in about the same proportions as their overall representation in the total population of Texas students.

However, **students who are overage and not on grade level constitute**

over 80 percent of the dropouts, almost 50 percent greater than their percentage of the total enrollment. This suggests that the majority of students who drop out are having academic difficulties in school. Although TAAS exit level test scores may confirm a lack of adequate academic progress for these students, they do not create it. Furthermore, the slightly lower percentage of Title I dropouts relative to the percentage of students enrolled in Title I programs suggests that remediation efforts have achieved some success in deterring dropouts.

Reasons for Leaving School Early. Examination of the reasons students leave school also indicates that TAAS exit level test performance plays a minor role in students' decision-making. The chief reasons reported by districts for 58 percent of the students leaving school in the 1996-97 school year are presented by ethnic group. These data indicate that the majority of students in all ethnic groups are leaving school due to academic difficulties related to poor attendance or low or failing grades. African-American students leave more often for alternative nondiploma programs (e.g., cosmetology school), while civilian or military employment is more attractive for Hispanic students.

Failing TAAS and not meeting all graduation requirements constituted a relatively small percentage, similar in magnitude to the percentage of students who were expelled for noncriminal behavior. About 2 percent of African-American and Hispanic students left school due to TAAS and graduation requirement deficiencies, while about 1 percent of whites left school for the same reason. However, even if these students had passed the TAAS exit level test, their incomplete high school records would have prevented them from receiving a high school diploma.

In addition to academic difficulties, peer pressure also plays a role in encouraging Hispanic students to leave school early. In some neighborhoods, it is considered "Anglo" and "nerdy" to do well in school. Also discouraging is the higher dropout rates for children of American-born Hispanics than for the children of immigrants, especially because the majority of Hispanic dropouts are American-born and fluent in English.[38]

Some Hispanics also believe that schools disrespect their culture and set academic expectations for Hispanic students too low. However, the TAAS exit level test does just the opposite; item content is carefully screened to eliminate offensive material, and the same standard of achievement applies to all students equally.

Dropout Recovery. For the 1996-97 dropout data compiled by TEA, students who had been reported as dropouts but whose whereabouts could be tracked were recovered back into the system. Categories of recovered dropouts included moving to another district, enrolling in an approved alternative program, returning to their home country, receiving a GED certificate, already having been counted as a dropout from another district in a previous year, being expelled and incarcerated for criminal behavior, with-drawing to attend college, exceeding the age limit for special education services, graduating, being reported more than once, or having met all graduation requirements except exit TAAS.[39]

Out of 217,533 students enrolled in grade 12, 1,782 students were identified as having met all graduation requirements except exit TAAS. If these students are included in the overall dropout calculations, the total percent of 12th-grade dropouts increases from 2.5 percent to 3.3 percent, a gain of 0.8 percent.

Based on these data, approximately eight out of 1,000 seniors met all the high school graduation requirements in their districts but failed to receive a high school diploma at their scheduled graduation in the spring because they had not yet passed the TAAS exit level test. These students had additional opportunities to pass TAAS at subsequent administrations or to obtain a GED equivalency certificate at a later date.

Comparing Dropouts and Remediation. The data provide a comparison of the magnitude of minority students successfully remediated relative to those leaving school due solely to failing the TAAS exit level test. Note that most of the students dropping out due to failing TAAS also failed to complete their high school requirements. Thus, the number of minority students meeting all graduation requirements except TAAS was relatively small.

For African-Americans, (.008)(30,801) = 246 failed TAAS but met all other high school graduation requirements. In contrast, about 13,362 African-American students passed the TAAS exit level test following remediation. For Hispanics, (.008)(69,038) = 552 seniors left school due solely to failure on the TAAS exit level test while 31,153 were successfully remediated.[40] Once again, the benefits of the TAAS exit level test in increasing the skill level of substantial numbers of minority students can be demonstrated to far outweigh the number of minority students discouraged by their test performance from completing the requirements for a high school diploma.

V. Predictors of TAAS Success

Common sense suggests that students who begin high school with adequate prerequisite skills and take more academic courses are more likely to pass the TAAS exit level test. Data collected by TEA support these relationships.

The percent of students passing the TAAS exit level mathematics subtest by course for a subset of the spring 1995 10th-grade students is reported for each ethnic group. For all ethnic groups, the passing rates increased for each higher level math course taken. Students receiving credit for Algebra II had the highest TAAS Mathematics passing rates while those receiving credit for Pre-algebra had the lowest passing rates. Minority students who received credit for Algebra II passed TAAS Mathematics at a rate four to five times higher than those receiving credit for Pre-algebra.

However, the data for the same spring 1995 10th grade students indicate that minority students took significantly fewer advanced math courses than white students. While the percentage of African-American and Hispanic students receiving credit for Algebra I was similar to that for white students, the percent of white students receiving credit for Geometry and Algebra II was, respectively, 1.5 and 2.0 times that for African-American and Hispanic students. In sum, these data indicate that minority students who complete advanced mathematics courses pass the TAAS exit level Mathematics subtest at much higher rates but that far fewer minority students than white students are completing advanced mathematics courses.

It is important to note two things about the data depicting a relationship between mathematics courses taken and TAAS exit level mathematics performance. First, advanced math courses are not required to pass the TAAS exit level Mathematics subtest. The math skills tested on the TAAS exit level mathematics subtest include content through eighth-grade math. The higher passing rates for students receiving credit for more advanced mathematics courses may be due to instructional reinforcement of prerequisite lower level content in the higher level courses. S*econd, Algebra I is now required for high school graduation in Texas.*

For the TAAS exit level Reading and Writing subtests, there is less variability in course taking because nearly all high school students are required to take English I and English II in their freshman and sophomore years. However, combined data from a set of case studies conducted by TEA indicate that grades received in English II are strongly related to TAAS exit level Reading performance.[41] The percent of students from the case study sample passing the 1995 TAAS exit level Reading subtest by course grade in English II indicate that about 90 percent of students earning A's and B's passed while only about 40 percent earning D's and F's did. The relationship between grades and TAAS performance is not perfect because courses may cover content different from TAAS, and grades may be based on factors other than achievement.

Finally, students who come to high school with adequate academic skills have a substantially higher likelihood of passing the TAAS exit level test. The percent of students passing all grade 8 TAAS tests in 1993 who passed all TAAS exit level tests in grade 10 is tabulated by ethnic group. As indicated, the TAAS exit level passing rates for students who passed TAAS in eighth grade are 80 percent to 90 percent. This relationship also is not perfect. Although attainment of eighth-grade skills indicates that students are ready to learn new material in high school, it does not guarantee that they will do so at a satisfactory level.

VI. The Shapiro Analysis

Dr. Shapiro's primary argument is that the TAAS test is "biased" because p-value differences between majority and minority groups correlate more highly with total group point-biserials than with minority group point-biserials. This argument is unfounded for four major reasons.

1."Bias" Measures Require Groups of Equal Achievement

First, p-value differences are significantly influenced by differences in ability between the two groups. As indicated earlier, comparisons designed to quantify bias must compare groups of equal ability. To the extent that p-value differences are based on groups of unequal ability, the purported measure of "bias" is confounded by achievement differences in the two groups.

For the 1994 and 1997 TAAS exit level tests analyzed in Dr. Shapiro's report, the achievement of African-Americans and Hispanics is below that of whites. For example, on the 1994 and 1997 TAAS exit level mathematics subtests, the mean p-values by group were as follows:[42]

	1994	**1997**
African-American	.57	.67
Hispanic	.63	.69
White	.76	.81

To construct a valid measure of performance differences due solely to "bias" attributable to an item, African-American, Hispanic, and white students with the same achievement level should be compared.

One method for measuring differential performance for groups of unequal achievement is to compare Rasch model item difficulties. Because all students responded to the same set of base items, item difficulties centered on zero for each group provide item measures that are not dependent on overall achievement level. That is, the mean Rasch item difficulty for each group is zero so differences in item difficulty between whites and African-Americans or whites and Hispanics measure a combination of estimation error and possible item "bias."

The data compare the Shapiro correlations of p-value differences with the values obtained using Rasch difficulty differences for the 1997 TAAS exit level Mathematics subtest. Note that the Rasch correlations are negative because large Rasch values denote hard items whereas large p-values indicate easy items.

The data indicate that when the effects of unequal achievement are removed, the correlations decrease substantially. In addition, the differential effects between correlations based on the total group point-biserials versus the minority group point-biserials also decrease.

2. P-value Differences and Point-biserials Measure the Same Item Characteristic

Second, it is not surprising to find that p-value differences for groups of unequal achievement and their corresponding total group point-biserials are positively correlated because the former is a crude measure of the latter. Point-biserials measure the degree to which persons who answer an item correctly tend to also have high total test scores and vice versa.

Another common method for quantifying the tendency for high-scoring students to answer a test item correctly and low-scoring students to answer incorrectly, D, is based on p-value differences between students with the highest test scores and those with the lowest test scores.[43] Generally these groups are formed using the upper and lower quartiles of the test distribution. However, given their unequal mean test scores, p-value differences between whites and African-Americans or whites and Hispanics will approximate the D statistic.

Two measures of the same characteristic will tend to rank order test items similarly and be moderately to highly correlated, depending on the accuracy of the measures. Thus, for the 1994 and 1997 TAAS exit level tests, the correlations between majority/minority p-value differences and total group point-biserials of .73 & .65 for African-Americans and .66 and .63 for Hispanics reported by Dr. Shapiro are in the expected range for alternative measures of the same characteristic.

3. Total and Minority Group Point-Biserial Distributions Are Similar

Third, the purpose for computing item point-biserials is to select items which students who have attained the tested skill answer correctly and those who have not attained the tested skill answer incorrectly. An item for which the reverse is true, that students with poor skills answer correctly while high-achieving students answer incorrectly generally has more than one correct answer or an ambiguity in wording that misleads the high-achieving students. Appropriate test development practice eliminates such items based on item point-biserials.

Using total group point-biserials for this purpose would be unfair to minority students only if their point-biserials tended to rank order the items differently, that is, if the same item flaw had differential effects for white students than for African-American and Hispanic students. The TAAS exit level data indicate that this is not the case. For example, for the 1997 TAAS exit level Mathematics subtest, the mean and standard deviation of point-biserials by group were as follows:[44]

	Mean	SD
African-American	.45	.06
Hispanic	.46	.06
Total	.47	.06

Clearly, the distributions of point-biserials in the two minority groups are very similar to the distribution of point-biserials for the total group.

4. Differences Between Highly Correlated Measures Are Unreliable

Fourth, the higher the correlation between the variables that constitute a difference score, the more unreliable is the difference score.[45] Item p-values and Rasch item difficulties are highly correlated for majority and minority groups. For example, for the 1997 TAAS exit level Mathematics subtest, the correlations are as follows:[46]

	White	
	p-value	**Rdiff**
African-American p-value	.93	
Hispanic p-value	.95	
African-American Rdiff		.95
Hispanic Rdiff		.97

Given these high intercorrelations, the reliabilities of differences in p-values or Rasch difficulties will be very low. That means that the differences are measuring primarily error. The correlation between an unreliable measure and another measure has little interpretive validity.

VII. The Haney Report

Dr. Haney comments on five major topics related to the use of TAAS as a graduation test. I disagree with his position in each of these areas for the following reasons.

1. Historical Use of Tests. The historical review of test use is interesting but lacks relevance to the TAAS exit level test. The TAAS exit level test is an achievement test, not an intelligence test. TAAS measures teachable academic skills that are clearly specified and disseminated.

Historical misuses of intelligence tests are unfortunate but have no bearing on the use of TAAS as a graduation test. Historical intelligence tests purported to measure an individual's fixed, innate ability. On the other hand, achievement tests measure learned academic content that is sensitive to instruction. Thus, achievement test scores are not fixed but change over time as students receive instruction and learn the tested skills.

The purpose of the TAAS exit level test is to identify those students, majority and minority, who have not yet attained the state exit level objectives in reading, mathematics and writing and to require remediation of the identified deficiencies. Although an unsuccessful first TAAS attempt can be discouraging, minority students would be even more disadvantaged if schools failed to identify and remediate their skill deficiencies. Such students would hold a diploma for seat time but would not have the academic skills expected of high school graduates in Texas.

Unlike the Florida students in the *Debra P.* case, African-American and Hispanic minority students subject to the TAAS exit level testing requirement have not been required by statute to attend segregated schools. As indicated in Dr. Haney's report, as soon as Florida high school students had all been educated in unitary schools and the state demonstrated the curricular validity of its graduation test, the courts upheld the use of the test to award diplomas. The *Debra P.* appeals court held:

> We affirm the district court's findings (1) that students were actually taught test skills, (2) that vestiges of past intentional segregation do not cause the [test's] disproportionate impact on blacks, and (3) that use of the [test] as a diploma sanction will help remedy the vestiges of past segregation. Therefore, the State of Florida may deny diplomas to students...[47]

2. Adverse Impact of Exit Level TAAS. As indicated in an earlier section of this report, differential performance between white and African-American and between white and Hispanic students meets the 80 percent standard for adverse impact for initial passing rates but not for cumulative passing rates. These data also indicate that the gap between passing rates for majority and minority students has narrowed so that initial minority passing rates are approaching the 80 percent standard. Further, TAAS exit level data demonstrate that large numbers of minority students are being successfully remediated.

The initial data set presented by Dr. Haney as evidence of adverse impact are misleading because these data are based on field-test information. Field-test data tend to exaggerate the number of failing scores because performance is lower when a test does not count. Thus, the data presented in Table 1 of Dr. Haney's report represent the worst possible case. The data set presented in Table 2 of the report (p. 11) is more appropriate and demonstrates that minority passing rates for the TAAS exit level test have increased substantially since the original TAAS field test with concomitant decreases in the gap between minority and majority group performance.

Dr. Haney's reference to an allegation that the TAAS exit level tests have become easier in the last few years (p. 10) is puzzling. These allegations are based on readability analyses of passages on the reading test. But as I expect Dr. Haney is aware, the difficulty of a reading test is reflected in the interaction between passages and items. Easy passages can be assessed with difficult items or hard passages with easier items. Therefore, readability analyses alone are not sufficient for judging the relative difficulty of different TAAS exit level forms; item data are also essential.

Because each new TAAS exit level test form is carefully developed to be parallel to previous forms and is equated to previous forms, the level of achievement required for passing remains constant. Equating adjustments have been extremely small in recent years indicating that the newly developed test forms are similar in difficulty to previous TAAS exit level test forms.[48] Finally, if the allegations Dr. Haney cites were true, it appears he would be arguing that the TAAS exit level test is both too easy and too hard for minority students.

As I indicated earlier in this report, statistical tests of adverse impact are unnecessary and inappropriate because the data being used are based on population values, not sample values. Even if such tests were appropriate, their results can be misleading when sample sizes are extremely large. That happens because errors become extremely small as samples become extremely large. When errors are extremely small, a statistical procedure can accurately infer small population differences from sample differences even when the population differences are too small to have any practical significance.

No statistical tests are required to ascertain that the Texas data reflect actual differences among ethnic groups. The relevant question is whether the observed differences are practically significant, and if so, whether they are caused by the TAAS exit level test or are the result of other factors. Application of the 80 percent standard is a reasonable way to answer the first part of the question. Dr. Haney answers the second part when he states: "[M]y own view . . . is that social, economic, and educational factors are the main determinants of the relative standing of ethnic groups on test results" (p. 7).

3. Grade Retention and Dropouts. Dr. Haney argues that schools are retaining in ninth grade students who are likely to fail the TAAS exit level test in 10th grade. He further argues that this is negative for minority students. However, districts are required to have specific guidelines for retaining students, and the typical deficiency of retained students is failure to earn sufficient course credits to qualify for sophomore standing. Alternatively, one could view this as a plus because it means that unprepared students are receiving additional instruction before attempting TAAS for the first time. This probably decreases frustration and increases the odds of passing.

If some schools are retaining students for the wrong reasons, this is not the fault of the test but of the human decision-makers who do so. Furthermore, a school cannot retain students indefinitely; to avoid affecting its accountability rating, the school must remediate unprepared students or they will be unsuccessful on TAAS in later years. Thus, any school engaging in such a practice has at worst temporarily delayed the day of reckoning and at best may have raised its scores for the following year by remediating unprepared students. Moreover, students retained in grade typically are also behind in their schoolwork, a condition the TAAS test may identify but does not cause.

As with passing rates, statistical tests for differences in retention rates are unnecessary because the data reflect population values.[49] The retention rates between majority and minority groups are clearly different, but so is the average achievement of these groups as measured by TAAS grade 8 scores and other nationally-normed standardized tests. One would expect more students to be retained in grade from groups demonstrating lower levels of achievement.

Texas dropout data were discussed earlier in this report. These data indicate that dropout rates have steadily declined over the past decade and that this trend was not affected by the introduction of the TAAS exit level test.

Dr. Haney poses a tough dilemma for schools: Is it better to hold unprepared students back and risk over-age dropouts or promote them to unsuitable coursework and almost certain failure on their first attempt to pass the TAAS exit level test? Unfortunately, this dilemma and the social conditions related to it would not abate if the TAAS exit level test were eliminated. High school graduates earn more than dropouts because they have higher skills, not more seat time. Promoting or giving diplomas to unprepared students would not solve the wage gap between skilled and unskilled workers nor would it eliminate the social problems experienced by unskilled workers.

4. Use of Exit Level TAAS in Isolation. Contrary to Dr. Haney's assertion, the TAAS exit level test is not used in isolation to make graduation decisions. In addition to passing the TAAS exit level test, students must successfully complete all required coursework and other graduation obligations imposed by their districts.

Dr. Haney proposes allowing high school grades to compensate for poor test performance. As indicated in my earlier discussion of this matter, although high school courses may cover the content and skills tested by the TAAS exit level test, teachers may grade students in part on nonacademic factors such as attitude, improvement or effort. The moderate correlations between course grades and TAAS exit level test scores cited by Dr. Haney indicate that TAAS tests and high school grades measure different student characteristics. This further supports the assertion that grades should not be viewed as substitute measures of tested content. Thus, it would be inappropriate to allow high grades to compensate for low scores on the TAAS exit level test.

For students who have not yet passed the TAAS exit level test by the date of their expected graduation, there are eight separate measures from eight different occasions indicating that they have not demonstrated satisfactory achievement of the state objectives. Moreover, it is virtually impossible for the true achievement of such students to be at or above the TAAS exit level passing standard.[50] Thus, these students are not false negatives and the decision not to award them high school diplomas is fully justified.

5. Lack of TAAS Validity Evidence. Earlier sections of this report provided a detailed analysis supporting the conclusion that the TAAS exit level test meets all relevant professional and legal standards. Specifically, the

TAAS exit level test is valid and reliable for its intended use as a graduation test.

Dr. Haney's statement that low correlations of TAAS exit level scores with grades and other measures indicate a lack of test validity is misleading. Correlations are an indicator of criterion-related validity which is appropriate for tests used to predict a criterion. For an achievement test, such as the TAAS exit level test, that directly measures specified state objectives, content validity evidence is most salient. The content validity evidence for the TAAS exit level test is extensive and convincing.

Similarly, Dr. Haney's assertion that the TAAS exit level test lacks curricular validity is also contradicted by the evidence. As discussed previously, all districts are required to teach the state-mandated curriculum; the state objectives, instructional targets, and released tests have been widely disseminated; state law requires districts to provide intensive remediation to students who fail the TAAS exit level test; study guides are being provided to parents of students who fail the test; and adequacy of preparation reviews by educator committees and bias review panels composed of minority members have demonstrated that the majority of Texas educators are teaching the tested content.

Dr. Haney notes in the introduction to his report that he has found TAAS useful for improving instruction in elementary and middle schools in Texas. These same positive TAAS qualities are also found in the exit level test and have resulted in an improved high school education for thousands of African-American and Hispanic students in Texas. The TAAS exit level test has also increased the value of a Texas high school diploma. If passing exit level TAAS were to be eliminated as a graduation requirement, these benefits would be lost. Only by placing responsibility jointly and concurrently on students and schools can Texas reach its goal of making the state essential elements a part of the education of all students who earn a high school diploma.

VIII. The Fassold Report

Mr. Fassold justly observes that diploma denial at a student's scheduled graduation based on not yet having passed the TAAS exit level test occurs only after multiple attempts spread across a two-year period. He states an intent to analyze cumulative passing rates across all TAAS exit level test administrations for the 1995 sophomore cohort scheduled to graduate in the spring of 1997. He then proceeds to present a series of single administration statistics.

Mr. Fassold based his analyses on several data files provided by TEA.

The data presented are inconsistent with the values posted on the TEA web page for all nonspecial education students passing all tests taken and the corresponding data reported in the TEA publication Student Performance Results, 1994-95 (p. 81) as shown below:[51]

	March '95 Column of Table D-1	**TEA Data Reports**
African-American	42.50	32%
Hispanic	44.83	37%
White	69.70	70%

The numbers in Table D-1 also do not match the passing rates when permutations of the following student characteristics are counted (or not counted) in the totals: special education students, nontested students, students who completed all three subtests, students who submitted an answer document, or students who passed a single subtest. Other possible reasons for the discrepancy include the inadvertent counting of students from other grades, students from other test administrations or other unspecified decisions that were made about which students to include in the numerator and denominator of the ratio.

Because it is unknown at this time why the Table D-1 data do not match the TEA reported data or any of the attempted replications completed so far using the data files provided to Mr. Fassold, the credibility of all of the data generated for his report is questionable.

I also question the methodology of the report for the following reasons:

(a) statistical tests appropriate for samples are reported for population differences.

(b) the designated "control" group lacks justification—if the reported data were correct, the control group data presented in Table D-2 would only indicate that passing rates for students without risk factors are generally higher than for the total group, and that being at risk is apparently not a satisfactory explanation for the differential performance among ethnic groups; in any case, such data do not establish the cause of the observed differences.

(c) only graduating seniors are eligible for the April/May TAAS exit level retest—therefore, the May '95 and May '96 administrations were not available to the spring 1995 10th-grade cohort, resulting in a total of eight attempts for this group, not 10 as listed in the report (p. 8).

(d) using the initial cohort size in the denominator of the "satisfaction rate" artificially depresses the values for minorities that have a higher percent of dropouts—it is not possible to determine from the data whether dropouts who had stayed in school and received remediation would have passed the TAAS exit level test.

(e) an unknown number of students who did not drop out and had not yet passed the TAAS exit level test may also not have passed all the courses required for graduation.

(f) the calculations described in the text for some of the tables are discrepant with the data given—for example, the calculations for satisfaction rates (p. 9) described in the text do not match the numbers given in Table F-2.

(g) if the data for LEP students are correct (p. 10), the cumulative passing rates for African-American, Hispanic and white LEP students were 47 percent, 45 percent and 52 percent, respectively, with the minority passing rates exceeding the 80 percent standard (80 percent of 52 percent = 42 percent).

(h) the school quality analysis (p. 10) is flawed because it assigns student quality measures based on the district the student attended—ratings for individual schools within a district can vary considerably from the district average making the latter statistic inaccurate for estimating school quality for individual students.

For example, in 1995, Houston ISD had three exemplary, seven acceptable and 17 low-performing high schools and had an overall district rating of accredited warned (academically unacceptable); Dallas ISD had one exemplary, 21 acceptable and four low-performing high schools with an overall district rating of accredited (academically acceptable).[52]

IX. Continuing Analysis

With less than a month available to study the plaintiffs' expert reports and to write this report, there was insufficient time to analyze all the data presented and to verify the results through replication. These activities are continuing and are expected to yield additional relevant information. Further, plaintiffs' experts have indicated that their reports are incomplete and have stated an intention for future supplementation. Thus, this report should be considered preliminary and subject to amendment.

X. Conclusion

The change from TEAMS to TAAS exit level testing changed the expectations for high school graduates from basic skills to higher-level academic skills. TAAS implementation has benefited Texans in the following ways:

- Increased the level of knowledge and skills attained by students earning a high school diploma;
- Increased the value of a high school diploma for minority and majority students;
- Identified and provided intensive remediation to unprepared students;
- Closed the gap between minority and majority performance and posted

cumulative passing rates for African-American and Hispanic groups that exceed the 80 percent adverse impact standard;

- Demonstrated minority gains on TAAS consistent with improvements on other standardized assessments;
- Focused attention on the educational needs of minority students who are not able to pass the TAAS exit level test on the first attempt.

Eliminating the TAAS exit level test—

- would remove a valid, reliable and fair measure of student achievement of state objectives;
- would probably not change the dropout rate appreciably;
- would not cause minority students to learn more;
- would remove important information in the accountability system for holding schools responsible for the achievement of all students;
- would remove the incentive for remediation that has narrowed the achievement gap between minority and majority students; and
- would reduce the value of a high school diploma in Texas.

Retaining the TAAS exit level test, but eliminating the requirement that students achieve a passing score to receive a high school diploma, would also compromise its benefits. Both schools and students must be held accountable for educational achievement to improve.

In summary, the TAAS exit level test is a high-quality testing instrument that meets all professional standards for large-scale achievement tests. Research to explore new technologies and improve the TAAS test instruments is an ongoing process. The TAAS exit level test did not create the social problems faced by minority groups but has contributed to their improvement. Passing the TAAS exit level test should be retained as a graduation requirement because its benefits to minority students far outweigh its alleged and unproven social costs.

Notes

[1] *Golden v. Birmingham Board of Education & Alabama State Department of Education*, IDEA due process hearing, January 1994; *Pandazides v. Virginia Board of Education*, 752 F. Supp. 696 (E.D. Va. 1990), *rev'd*, 946 F.2d 345 (4th Cir. 1991).

[2] American Psychological Association (1985). AERA/APA/NCME *Standards for Educational and Psychological Testing*, Washington, D.C.

[3] The *Debra P.* case involved two separate trials, four years apart. The initial case was brought in 1979 in Federal District Court and was appealed, *Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979), *aff'd, in part, rev'd, in part*, 644 F.2d 397 (5th Cir. 1981). The case was remanded for further proceedings, including a determination of whether the Florida test met the notice and curricular validity requirements imposed for the first time by the appeals court. The second trial occurred in

1983 and was also appealed, Debra P. v. Turlington, 564 F.Supp. 177 (M.D. Fla. 1983), *aff'd,* 730 F.2d 1405 (11th Cir. 1984). The court found that the Florida test met the notice and curricular validity requirements and allowed Florida to award diplomas based on the test results.

[4] The TAAS exit level objectives are a subset of the critical components of the state-mandated curriculum.

[5] See the *1997 Technical Digest* Appendix 5 for a description of item review considerations used by the educator committees.

[6] *Texas Student Assessment Program Technical Digest 1996-97* [1997 *Technical Digest*], p. 5, 8.

[7] Selected State AEIS Data, Five Year History, www.tea.state.tx.us, Nov. 17, 1998.

[8] Due to time constraints, written composition essays are field-tested separately.

[9] *1997 Technical Digest,* pp. 277-79.

[10] *1997 Technical Digest,* p. 43.

[11] Memo from NCS Psychometric Services, December 1998; *1997 Technical Digest,* p. 273.

[12] A student may also continue to retake the TAAS exit level test at each administration following the student's scheduled graduation. If successful, and if all other requirements for graduation have been met, the student is awarded a high school diploma. *TEC* § 39.025 (b), (c).

[13] Chapter 3, p. 25-30; Chapter 5, pp. 35-37.

[14] *1997 Technical Digest,* Chapter 2 & Appendix 3 pp. 147-151.

[15] § 101.4 Test Security and Confidentiality, (a) - (d).

[16] *Texas Education Code,* § 39.030.

[17] *1997 Technical Digest,* Chapter 5, pp. 20-25.

[18] § 39.024 (a).

[19] Minutes State Board of Education Meeting, July 1990, *1997 Technical Digest,* Appendix 9, pp. 352-54.

[20] NCS, TPC, & MI, *TEAMS and TAAS Annual Report,* 1989-90.

[21] § 101.2 (e).

[22] Similarly, students who were first administered the TAAS exit level test under the 60 percent phase-in standard continue to have that standard applied on subsequent attempts.

[23] *Standard* 8.7, p. 53.

[24] *Debra P. v. Turlington,* 564 F. Supp. 177, 186 (M.D. Fla. 1983).

[25] *Debra P.* at 183-86.

[26] *1997 Technical Digest,* Appendix 5, p. 164.

[27] *Debra P. v. Turlington,* 730 F.2d 1405, 1410-11 (11th Cir. 1994).

[28] *TEC* § 39.023 (e).

[29] Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1985).

[30] In May 1998, TEA reported that 96.7 percent of the Class of 1998 had passed the TAAS exit level test. *Growing Number of Seniors Pass TAAS Test,* TEA Press Release, May 27, 1998, @www.tea.state.tx.us, Dec. 23, 1998.

[31] *Debra P. v Turlington,* 654 F.2d 1079, 1088 (5th Cir. 1981).

[32] Data from 1998 indicate that average SAT I scores for African-American and

Hispanic students remained stable or slightly increased, despite a record increase in the number of students taking the test. *Number of Texas students taking SAT and AP exams rises substantially,* TEA Press Release, Sept. 1, 1998, @www.tea.state.tx.us, Dec. 23, 1998.

[33] *Successful Texas Schoolwide Programs: Research Study Results (February 1997),* @www.starcenter.org, Dec. 23, 1998.

[34] *Eleven High-Achieving, High-Poverty Schools Asked to Participate in Successful Schools Study,* TEA Press Release, May 13, 1998, @www.tea.state.tx.us, Dec. 23, 1998.

[35] Headden, S., The Hispanic Dropout Mystery, *U.S. News & World Report,* Oct. 20, 1997, p. 64.

[36] Headden, S., The Hispanic Dropout Mystery, *U.S. News & World Report,* Oct. 20, 1997, p. 64.

[37] 1994-95/1995-96 Public Elementary/Secondary School Agency Universe Survey, National Center for Education Statistics, December 1997, p. 17.

[38] Headden, S., The Hispanic Dropout Mystery, *U.S. News & World Report,* Oct. 20, 1997, pp. 64-5.

[39] 1994-95/1995-96 Public Elementary/Secondary School Agency Universe Survey, National Center for Education Statistics, December 1997, p. 5.

[40] Chart 8; 1994-95/1995-96 Public Elementary/Secondary School Agency Universe Survey, National Center for Education Statistics, December 1997, pp. 5, 7, 11.

[41] TEA, *Student Performance Results 1994-95,* pp. 189-196.

[42] *1997 Technical Digest,* p. 277-79; NCS Statistical Analysis, 12/28/98.

[43] Crocker, L. & Algina, J. *Introduction to Classical and Modern Test Theory,* New York: Holt, 1986, pp. 314-15.

[44] NCS Statistical Analysis, 12/28/98.

[45] Crocker, L. & Algina, J. *Introduction to Classical and Modern Test Theory,* New York: Holt, 1986, p. 150.

[46] NCS Statistical Analysis, Dec. 28, 1998.

[47] *Debra P. v. Turlington,* 730 F.2d 1405, 1416-17 (11th Cir. 1984).

[48] Memo from NCS Psychometric Services, December 1998.

[49] Note that the *Castaneda v. Partida* case cited by Dr. Haney involved the selection of *samples* of jurors from a regional adult population. Statistical tests were appropriate in this case for evaluating the credibility of the proportional group representation in a jury sample purported to have been randomly selected from the regional adult population.

[50] The probability of failing the TAAS exit level test eight times given a true achievement at or 1 SEM above the passing standard = .0001 to .004.

[51] Incidentally, the passing rate reported for Native Americans in Table D-1 for March 1995, 83 percent, is even more discrepant from the TEA reported value of 52 percent.

[52] *1998 District Accountability Rating History,* @www.tea. state.tx.us, Dec. 23, 1998.

Expert Report of William A. Mehrens

January 1999

Background

My name is Dr. William A. Mehrens. I am a professor of educational measurement at Michigan State University. My address is 462 Erickson Hall, Michigan State University, East Lansing, MI 48824. I have worked professionally in the field of educational and psychological measurement since 1965. During that time I have conducted research, published textbooks and articles, taught and advised graduate students, consulted, served as an expert witness, and served in several elected positions for various professional organizations.

To provide a bit more detail on my professional background, I received a bachelor of science degree in 1958 with dual majors in mathematics and chemistry from the University of Nebraska. I received a master's of education in 1959 with a major in educational psychology also from the University of Nebraska. I received a Ph.D. in 1965 from the University of Minnesota. My major was educational psychology with an emphasis area in measurement. I taught mathematics in a public junior high school and was a counselor in a public high school in Minneapolis. I am a member of several professional organizations and have held elective office in several including, but not limited to, the American Educational Research Association (previously secretary and vice president of Division D) and the National Council on Measurement in Education (previously served on the board of directors and as president).

I have co-authored a number of textbooks. One that may be most relevant is entitled *Measurement and Evaluation in Education and Psychology.* It is currently in its fourth edition. I have published over 80 articles or book chapters, written over 100 reports, and have presented over 180 major speeches. My vita, which includes cases where I have testified or been deposed, has been submitted previously.

I have been asked to review various documents related to the Texas Assessment of Academic Skills (TAAS) exit test and present my professional opinions regarding the degree to which the test meets the professional standards in the field of educational measurement. The opinions expressed herein are my own.

Documents Reviewed

As of this point in forming my opinion I have reviewed the FIRST AMENDED COMPLAINT (September 1, 1998); the *Texas Student Assessment Program: Technical Digest for the Academic Year 1996-1997* (hereafter called the *Technical Digest*); a draft report by Jaeger and Busch on the Review of the Texas Assessment of Academic Skills (it is my understanding a final report was never issued); a *Response* to a paper entitled TAAS and Accountability: Review, Analysis, and Policy Implications prepared by National Computer Systems, The Psychological Corporation and Measurement Incorporated (dated October 28, 1994) hereafter just called the *Response*); a memo from Susan Phillips to Elliott Johnson dated September 6, 1994; a memo from Twing to Phillips dated December 13, 1998; and expert witness reports for the plaintiffs prepared by Bernal, Cardenas, Fassold, Haney, McNeil, Shapiro, Valencia, and Valenzuela.

Overview of My Opinions

I have formed several opinions that are stated here in outline form. The substance, bases, and reasons for these positions are all stated more fully within subsequent sections of this report.

(1) The TAAS has been constructed according to acceptable professional standards.

(2) The TAAS tests curricular material that the state views as important for graduates to have mastered.

(3) Without a requirement like the TAAS, students might graduate without having learned what the state has deemed to be a set of minimal requirements.

(4) Students have had ample opportunity to learn the material tested on the TAAS.

(5) Providing instruction over the objectives tested by the TAAS seems commendable, not something to be condemned.

(6) Having a required exit examination like the TAAS should increase efforts to educate those subgroups of students who have, historically, not received an adequate education. Focusing remediation on those students who fail should assist in removing any alleged vestiges of discrimination in education.

(7) Requiring a test such as the TAAS should encourage schools to teach toward the objectives that the state has deemed appropriate for education. This seems preferable and certainly less discriminating than adjusting the content of the test to any perceived currently inadequate curriculum.

(8) The test is sufficiently reliable. Any unreliability works to the benefit of the examinees who have true scores below the actual standard because they receive eight opportunities to take the test and may eventually pass because of positive random errors of measurement.

(9) Allowing students eight opportunities to pass the exit level TAAS prior to graduation ensures that the probability of not passing due to random error is almost zero. That is, students will not fail the test eight times due to random error. Furthermore, allowing eight opportunities means that some students who have actual levels of achievement below the standard will pass due to positive random errors of measurement.

(10) Appropriate steps have been taken in the test construction process to ensure that the inferences intended to be drawn from the test scores are appropriately valid.

(11) Appropriate steps have been taken to minimize any potential bias in the test.

(12) Any "adverse impact" data should be based on the cumulative pass rate and should not be analyzed by an inferential statistics procedure.

(13) Conjunctive decision-making is an appropriate decision-making model. Using test data in this model is appropriate.

(14) Standard setting is a judgmental process. Those in authority should make this judgment. It was appropriate for the State Board of Education to set the cut score. They had sufficient information when they set the cut score.

TAAS Test Development

(1) The TAAS has been constructed according to acceptable professional standards.

The determination of what is an acceptable standard in test development must be based on professional judgment. It is certainly possible for measurement experts to have different opinions regarding how close to perfect a test must be to reach "professional standards." I feel very strongly that the question of appropriateness is not whether scholars hired to find fault with a process can succeed in finding fault. Any scholar in the field, no matter what test he/she looked at, could find ways to improve or criticize the test construction and validation process. We cannot hold to idealized, ivory tower standards, because if we did, no test would ever meet the standards. I reference quotes from the professional *Standards for Educational and Psychological Testing* (hereafter called the *Standards*) and its predecessor supporting the acceptability of non-perfect procedures. As stated in the most recent edition of the *Standards:*

> Evaluating the acceptability of a test or test application does not rest on the

> literal satisfaction of every primary standard in this document, and acceptability cannot be determined by using a checklist (AERA/APA/NCME, 1985, p.2).

As the previous edition of the *Standards* pointed out:

> "The individual standards are statements of ideals or goals..." (APA/AERA/NCME, 1974, p. 4, emphasis added).

Thus, I wish to stress that my professional judgment is that the TAAS meets a *reasonable standard of acceptability*, but that I do not claim that either the test, or its documentation (or any other test), could not be improved.

In constructing an appropriate achievement test to be used for high school graduation requirements, there are several basic steps that need to be taken: The content domain of the test must be determined; test specifications must be developed; items must be written, field tested, and evaluated; there should be item sensitivity/differential item functioning analyses; and a cut score must be set. All of these steps have been followed in an acceptable manner.

The basic test development procedures have been described in Chapter 2 of the *Technical Digest*. In 1988 and 1989 the Texas Education Agency (TEA) held meetings with more than 1,000 educators and concerned citizens. TEA staff members and advisory committees worked jointly to develop the objectives on which the TAAS was based. Test specifications and test blueprints were developed; items were developed and reviewed; they were pilot tested; items were field tested and the data from the field tests were used to build the final test; and differential item functioning statistics were used to determine whether any items were functioning differently across ethnic and gender groups. Page 13 of the *Technical Digest* presents a flow chart of the item development process. That flow chart, plus the accompanying text, documents that the test was developed in a careful, professionally acceptable fashion. As Jaeger and Busch stated in their review:

> the procedures used by The Psychological Corporation and National Computer Systems to assemble the TAAS tests appear to be reasonable and generally consistent with accepted psychometric practice (1994, p. 7).

The few minor concerns that report specified were concerns based on incomplete information available to them when they wrote the report (see the *Response*).

Appropriate Content

(2) The TAAS tests curricular material that the state views as important for graduates to have mastered.

The TAAS tests are based on the Texas essential elements which have been outlined in the State Board of Education Rules of Curriculum. Texas educators were involved in examining these essential elements when constructing the objectives on which the TAAS would be built.

(3) Without a requirement like the TAAS, students might graduate without having learned what the state has deemed to be a set of minimal requirements.

States implement high school graduation tests precisely because they wish to ensure that students have acquired a certain amount of knowledge and skills. Without such testing requirements, students could be given a diploma without learning these things. As Jaeger and Busch pointed out in their report, between 27 percent and 50 percent of students who fail the TAAS tests nonetheless pass corresponding courses. They suggest two possible reasons: grade inflation and opportunity to learn the content of the TAAS in their courses. From what I have read, the opportunity to learn hypothesis is not tenable (to be discussed more later).

Opportunity To Learn Issues

(4) Students have had ample opportunity to learn the material tested on the TAAS.

The schools in Texas are required by law to teach the essential elements outlined in the State Board of Education Rules of Curriculum. The TAAS has been built on the content in those essential elements. Furthermore, all students failing to attain the passing standard must be offered remedial instruction. Furthermore, applicants who fail have eight opportunities to take the test, with remediation opportunities between each assessment.

McNeil, in her expert report, opines that "TAAS drills are becoming the curriculum in our poorest schools" (p. 3). Later she suggests that "The press to spend instructional dollars on test-prep materials is widespread, especially among those schools with poor and minority children..." (p. 6). Whether or not teaching the content domain which the TAAS samples is a good or bad thing for a student could, I suppose, be debated. However, it is true that (a) Texas officials believe the content domain is important and (b) teaching the domain should increase students' levels of knowledge and skill on that content. (It should be pointed out that one cannot teach the specific questions on the TAAS because they are not known in advance. To improve on the TAAS, one has to learn more of the content domain that is sampled.)

At any rate, to the extent that the TAAS content is taken from the essential elements, that state law requires schools to teach these elements, that schools must provide remediation to students who fail, that there are eight opportunities to take the test, and that there is at least some opinion by experts for the plaintiffs that schools emphasize the content tested by the TAAS, it seems obvious that students do have ample opportunity to learn the material.

(5) Providing instruction over the objectives tested by the TAAS seems commendable, not something to be condemned.

Certainly it is possible to raise scores on non-secure tests without increasing the students' knowledge and skills on the domain which the test samples. When scores on tests go up, but knowledge of the domain does not, one could speak of test score pollution. I have written about this issue (see, for example, Mehrens and Kaminski, 1989). One can teach too closely to a test—especially if the items are not secure. For example, on a non-secure standardized test where the questions do not change for several years, schools could teach the actual questions. Since the hoped for inference is to a larger domain of knowledge which the test only samples, the teaching to the specific questions likely leads to an incorrect inference. When the test questions themselves are secure (i.e., not known in advance of the test being administered), it is not possible to teach the specific questions. What is possible is to emphasize in instruction the specific domain of content which the test questions sample. If the domain is an important domain (and Texas officials apparently think it is), it seems beneficial to increase instruction over that domain. Of course, one should not then infer that an increase in scores on the domain tested indicates an increase in the knowledge of some different domain.

McNeil opines in her expert witness report for the plaintiffs that: "successful performance on the TAAS in no way insures a quality education" (p. 9). Depending on one's definition of a quality education, this is likely true. That is why the TAAS is but one of the requirements for a high school diploma. Students must meet additional requirements as well. One should not make incorrect inferences from successful performance on the TAAS. However, successful performance on the TAAS does suggest that students have acquired a minimum amount of the knowledge and skills Texas officials have deemed to be important.

(6) Having a required exit examination like the TAAS should increase efforts to educate those subgroups of students who have, historically, not received an adequate education. Focusing remediation on those students who fail should assist in removing any alleged vestiges of discrimination in education.

It has been suggested by some experts for the plaintiffs that education in Texas has not always been distributed equally across subpopulations. To the extent this is true, the requirement of a test such as the TAAS along with the state requirement that schools must offer remediation to those who fail should help to ameliorate any alleged vestiges of discrimination.

(7) Requiring a test such as the TAAS should encourage schools to teach toward the objectives that the state has deemed appropriate for education. This seems preferable and certainly less discriminating than adjust-

ing the content of the test to any perceived currently inadequate curriculum.

Some individuals might wish for less standardization in a state test. Why not, they might argue, adjust the content to the local curriculum? The disadvantage of this approach would be to perpetuate inadequate curriculums in those schools.

Reliability Issues

(8) The test is sufficiently reliable. Any unreliability works to the benefit of the examinees who have true scores below the actual standard because they receive eight opportunities to take the test and may eventually pass because of positive random errors of measurement.

As pointed out in Chapter 8 of the *Technical Digest*, internal consistency reliabilities range from the high 80s to the low 90s (see page 41). These are acceptably high. The score reliability for the written composition shows acceptably high agreement rates (98.06 percent agreement rate for three readings—see p. 43). In addition, "all exit level compositions that receive a score of "1" undergo an extra round of scoring by a select group of specialists who have been trained exclusively on the "1/2" line" (p. 43).

The Jaeger and Busch report suggested that reliabilities and standard errors of measurement should be reported at the cut score. These suggestions are supported by the *Standards*. But, there is little practical importance in reporting reliabilities at the cut score. As Brennan has pointed out, an index of dependability for domain-referenced mastery interpretations can be no less than KR-21 when items are scored dichotomously (the case for most of the TAAS tests) (Brennan, 1984). KR-21 is only a slightly lower estimate than K-R 20, which is what is reported in the *Technical Digest*. In a memo from Twing to Phillips (1998), the standard errors at the cut score are reported. These range from around 0.29 in mathematics to 0.37 in Writing in terms of the Rasch ability scale. As he reports, these values represent raw score values from about 2½ to 3½ raw score points.

While reliability is acceptably high, it should be pointed out any unreliability helps examinees who have true scores below the actual standard because such candidates receive eight opportunities to take the test and may eventually pass because of positive random errors of measurement.

(9) Allowing students eight opportunities to pass the exit level TAAS prior to graduation ensures that the probability of not passing due to random error is almost zero. That is, students will not fail the test eight times due to random error. Furthermore, allowing eight opportunities means that some students who have actual levels of achievement below the standard will pass due to positive random errors of measurement.

As is pointed out on page 31 of the *Technical Digest*, an individual with a true achievement level of 70 has a 99.6 percent probability of passing after eight attempts. In fact, students with a true score of 62 (considerably below the passing standard) have a 55 percent probability of passing after eight attempts.

Validity Issues

(10) Appropriate steps have been taken in the test construction process to ensure that the inferences intended to be drawn from the test scores are appropriately valid.

Some of the current literature on the meaning of validity suggests that all validity evidence is, at bottom, construct validity evidence. As the *Standards* suggest: "evidence identified usually with the criterion-related or content-related categories...is relevant also to the construct-related category" (AERA/APA/NCME, 1985, p. 9). Nevertheless, the traditional divisions of content, criterion-related, and construct validity evidences exist in the *Standards* and in court precedents. For high school graduation tests such as the TAAS, the major evidence should be content validity evidence.

As correctly pointed out in the *Technical Digest,* criterion-referenced achievement tests such as the TAAS are based on an extensive definition of the content that they assess. The TAAS is "content-based and tied directly to the Texas essential elements, the state-mandated curriculum in place during the 1996-1997 school year" (p. 45). The *Digest* describes the steps taken to ensure that the TAAS test objectives are tied to the essential elements and that the items align with the objectives. With respect to the TAAS, "the construct tested is the mastery of academic content required by the state-mandated curriculum, in this case, the Texas essential elements" (p. 46). Thus, "the construct validity is grounded in the content validity of the test" (p. 46).

The intended inference to be drawn from TAAS scores is simply either that the individual test taker has, or does not have, a sufficient demonstrated minimum level of the competencies that the test is attempting to measure. It is a valid inference if the test reasonably measures the competencies in question. The TAAS certainly has been constructed to ensure that it, in fact, measures those competencies.

Bias Issues

(11) Appropriate steps have been taken to minimize any potential bias in the test.

Tests should be free of bias. However, it should be noted that differential performance across groups of individuals is not evidence of bias. It is important to discuss this issue more fully because an expert witness for the plaintiffs, Shapiro, indicates a total misunderstanding of item bias.

There has been much confusion about this issue among non-measurement specialists, probably exacerbated by a settlement between the *Golden Rule Insurance Company* and the *Educational Testing Service* typically referred to as the *Golden Rule*. That settlement required looking at the actual p-value (proportion correct) differences between blacks and whites in choosing items for a test. Following the *Golden Rule* agreement, there was considerable interest among measurement professionals regarding the impact of the agreement with respect to building a valid test. Many professionals thought carefully about the issue, did research, and wrote articles based on theoretical and logical arguments as well as the empirical evidence from their research. Those professionals viewed the *Golden Rule* agreement as faulty with respect to enabling professionals to build valid tests. Perhaps the quickest way to get an overview of the professionals' views is to simply abstract and/or quote from some of the many papers and articles that have been produced since the *Golden Rule* settlement. Dr. Bond, a very well known African American measurement expert makes the following points:

> It is axiomatic in psychometric circles that group differences in total test performance, *per se*, cannot be taken as evidence of test bias (Bond, 1987, pp 19-20).

An American Psychological Association Committee on Psychological Tests and Assessment (CPTA) has written as follows:

> First, the mere existence of differences between groups is not an accurate indication of bias...Differences between groups also may reflect valid behavioral differences...If group differences in knowledge or ability, as well as spurious differences arising from irrelevant sources of variance, are reflected in item statistics, a procedure is needed to distinguish the two. Well established procedures are available for this purpose....procedures that require choosing items on the basis of considerations other than those leading to optimal measurement of the relevant construct, are likely to lower the psychometric quality of the test (1988, pp. 4-5).

More scholars could be quoted, but I simply refer to such articles as the following (and quote a portion of one of them): Jaeger, 1987; Linn & Drasgow, 1987; Marco, 1988; Plake, 1995; and Shepard, 1987. All the referenced articles are by very well known and respected measurement specialists. They all take the position that looking at p-value differences (as Shapiro did) is simply an incorrect way to judge item bias. As Shepard has stated:

> I claim to be an advocate for the discovery of test bias. I am, however, strongly opposed to the Golden Rule procedure. [Mehrens note: She is referring to the p-value difference like Shapiro used.] The Golden Rule, despite its unfortunately benevolent name, will harm valid test construction and will undermine legitimate efforts to screen tests for bias....two essential points must be comprehended: 1) group differences in item passing rates are not indicators of bias, and 2) using passing rates in this way will lead to the selec-

> tion of the worst set of test items, i.e., those questions that are less reliable and more influenced by guessing (Shepard, 1987, p. 7).

The TAAS has received admirable attention related to the issue of bias. As the Jaeger and Busch draft report states: "It appears that substantial attention has been paid to the review of items on the TAAS tests to determine the extent and nature of item bias" (p. 18). The *Technical Digest* presents information on how item review committees looked at items for potential bias and describes the statistical procedures employed to look at differential item functioning. Items that were flagged statistically were reviewed, which is exactly how they should have been treated. As Plake has pointed out:

> It is important to note that differential item performance, per se, is not prima facie evidence that the item is biased (1995, p. 207).
>
> First and foremost, any item that shows differential item functioning must be scrutinized for bias. If differential performance is supported by the construct being assessed, then the differential performance is valid, and the item should be maintained in the operational test score (1995, p. 213—note here that Plake is talking about a DIF procedure, not just looking at mean differences).

Adverse Impact

(12) Any "adverse impact" data should be based on the cumulative pass rate and should not be analyzed by an inferential statistics procedure.

While I have not done any adverse impact analysis and do not intend to testify about any actual data on adverse impact, I may testify about how such analyses should be done. It is my opinion that the only reasonable data source would be the cumulative pass rate of the various groups. If an individual fails the TAAS on the first attempt, he/she has seven more attempts prior to the scheduled graduation date and there is an obligation that the student receive remediation. Surely it should not be considered harmful to have additional attention paid to one's learning of essential material. Therefore, for example, most of the analyses by Fassold and the analyses done by Haney are, in my opinion, irrelevant to the issue of adverse impact. The question is: What percent of the individuals passed after all their attempts?

In addition, in my opinion it is inappropriate to use any inferential statistics test when dealing with a population—particularly a population as large as high school students in Texas. The correct statistic to use is the 80 percent rule. (See Meier, Sacks, & Zabell, 1984.)

Conjunctive Decision-making

(13) Conjunctive decision-making is an appropriate decision-making model. Using test data in this model is appropriate.

A statement that has been used at times by some individuals to criticize the use of tests for high school graduation decisions is that decisions should not be based on only a single piece of information. The psychometric issue should not be whether more data lead to better decisions than fewer data. They do. And using the TAAS in addition to previously existing criteria for making a high school graduation decision is obviously using more data. The psychometric issue is how we should combine data. Possible methods include the conjunctive model and the compensatory model. When one uses a conjunctive model, an individual must score above the cut off on each of the measures used. In the compensatory model, high scores on one variable can compensate for low scores on other variables. Both of these models are appropriate under certain circumstances. They do not differ, as models, with respect to the amount, or type, of data that is gathered. They differ with respect to how one *combines* the various pieces of data to make a decision. There can be legitimate differences in opinion regarding which method produces the "better" decisions. In the extant case, I vote for the conjunctive method because I believe it is in the best interests of the students and society to have some minimal level of competence in Mathematics, Reading, and Writing. But whatever position one takes, Texas is not using only one piece of information. [The state is] using a conjunctive decision-making model.

Several experts for the plaintiffs have commented on using the TAAS as the sole criterion for making a decision about high school graduation. Several points need to be made with respect to this issue. First, note the quote from the *Standards* by Haney. Standard 8.12 states that a decision that will have a major impact on a test taker should "not automatically be made on the basis of a single test score" (p.54 of the *Standards* and p. 18 from Haney's Preliminary Report). This standard is not relevant because the decision made on the basis of a *single score* from the TAAS does not have a major impact. The only impact is that, if the test taker fails, he/she is provided with remediation. A more relevant quote would be Standard 8.8:

> Students who must demonstrate mastery of certain skills or knowledge before being promoted or granted a diploma should have multiple opportunities to demonstrate the skills (AERA/APA/NCME, 1985, p. 53).

That standard, which is the relevant one (and assumes requiring mastery prior to granting a diploma is an acceptable thing to do), is followed by Texas, which provides eight opportunities before the scheduled date of graduation and continues to provide opportunities after that date.

Cardenas suggested that "the common use of the term 'sole criterion' in educational literature denotes any criterion as 'sole' if it is used in determining a decision regardless of what other criteria must be met" (p. 14). This is simply not true. Why would one suppose it would be common to use "sole" if other criteria must be met?

In point of fact, to graduate in Texas one is given many opportunities to pass the TAAS, and passing of the test is NOT the only criterion for receiving a high school diploma. A student also has to complete course work. Bernal, in his expert report for the plaintiffs, seems to be arguing for a compensatory model across the three subject matter tests. He posits that "In our common experience successful high school graduates use their areas of greater skill to compensate for areas of relative weakness" (p. 2). However, what Bernal does not seem to realize in this analogy is that in order to receive credit for a high school course, you must pass it. You cannot use high achievement in an English course to compensate for a failing grade in a math course. High school graduation requirements that consist of passing so many courses are, in fact, conjunctive, not compensatory. So, contrary to what Bernal thinks is common experience, it is, in fact, common to employ the conjunctive model.

Standard Setting

(14) Standard setting is a judgmental process. Those in authority should make this judgment. It was appropriate for the State Board of Education to set the cut score. It had sufficient information when it set the cut score.

Chapter 6 and Appendix 9 of the *Technical Digest* present some information on standard setting (setting the cut score). As is pointed out in Chapter 6, "Texas law authorizes the State Board of Education to establish standards for the statewide assessment instruments. The Texas Education Agency supplies SBOE members with a wealth of data to help inform their decisions" (p. 28). Appendix 9 presents some, but apparently not all, of the information given to the SBOE. Included in the information were projected impact data showing the projected percent passing by black, Hispanic, white, and total. SBOE minutes show that they unanimously voted to approve the commissioner's recommendations regarding the standards.

There are a variety of methods used in the profession to set standards, and there is not a consensus about which method is best. In discussing the development of the *Standards*, Linn stated that:

> there was not a sufficient degree of consensus on this issue...to justify a specific standard on cut scores (Linn, 1984, p. 12).

The only Standard directly related to the methodology of setting a cut score is Standard 6.9, which provides that:

> the method and rationale for setting that cut score, including any technical analyses, should be presented in a manual or report. When cut scores are based primarily on professional judgment, the qualifications of the judges also should be documented (AERA/APA/NCME, 1985, p. 43).

While the *Standards* are more than a decade old, there is still not consensus within the profession regarding standard setting methodology. With respect to Standard 6.9, I would have wished for a bit more information in the *Technical Digest* regarding how the TEA developed its recommendation. However, it is clear that the cut scores were eventually based on the professional judgment of the SBOE, and its qualifications were a matter of public record.

There will always be individuals who will wish the cut score were set at a different point. However, it is well recognized that, because setting standards is a judgmental process, there is no right answer.

> A great amount of early work on standard setting was based on the often unstated assumption that determining a test standard parallels estimation of a population parameter—there is a right answer and it is the task of standard setting to find it. ... A right answer to the standard-setting question does not exist, except perhaps in the minds of those providing judgments (Jaeger, 1986, p. 195).

Conclusion

Tests must be judged against reasonable standards. The TAAS has been constructed in a professionally acceptable manner.

[This report was first published in Applied Measurement in Education, *vol. 13, no. 4, October 2000, and appears here by permission of the author and Lawrence Erlbaum, Publishers.]*

References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (AERA/APA/NCME), (1985). *Standards for educational and psychological testing.* Washington, D.C.: American Psychological Association.

American Psychological Association, American Educational Research Association, and the National Council on Measurement in Education (APA/AERA/NCME), (1974). *Standards for educational and psychological tests.* Washington, D.C.: American Psychological Association.

Bernal, E.M. (Undated, but apparently submitted in 1998). *Report of Dr. Ernesto M. Bernal.*

Bond, L. (1987). The Golden Rule settlement: A minority perspective. *Educational Measurement: Issues and Practice,* 6(2), 18-20.

Brennan, R.L. (1984). Estimating the dependability of the scores. In R.A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 292-334). Baltimore: The Johns Hopkins University Press.

Cardenas, J.A. (Undated, but apparently 1998). *Testimony: G.I. Forum, et al. v. TEA.*

Committee on Psychological Tests and Assessment (CPTA).(1988, August 11). *Implications for test fairness of the "Golden Rule" company settlement.* Washington D.C.: American Psychological Association.

Fassold, M. (1998, November 15). *Written report of Mark Fassold.*

Haney, W.M. (1998, December 11). *Preliminary report on Texas Assessment of Academic Skills Exit Test (TAAS-X).*

Jaeger, R.M. (1986). Policy issues in standard setting for professional licensing tests. In W. P. Gorth & M.L. Chernoff (Eds.), *Testing for teacher certification* (pp. 185-199). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jaeger, R.M. (1987). NCME opposition to proposed golden rule legislation. *Educational Measurement: Issues and Practice,* 6(2), 21-22.

Jaeger, R.M. & Busch, J.C. (1994, June 21). *Draft Report: Review of the Texas Assessment of Academic Skills.*

Linn, R.L. (1984, April). *Standards for validity in licensure testing.* Paper presented at the "Validity in Licensure Testing: symposium" at the annual meeting of the American Educational Research Association, New Orleans, La.

Linn, R.L., & Drasgow, F.D. (1987). Implications of the Golden Rule settlement for test construction. *Educational Measurement: Issues and Practice,* 6(2), 13-17.

Marco, G.L. (1988). Does the use of test assembly procedures proposed in legislation make any difference in test properties and in the test performance of black and white test takers? *Applied Measurement in Education,* 1(2), 109-133.

McNeil, L.M. (undated, but apparently 1998). *Expert witness for the plaintiff.*

Mehrens, W.A. & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless or fraudulent? *Educational Measurement: Issues and Practice,* 8(1), 14-22.

Meier, P., Sacks, J., & Zabell, S.L. (1984). What happened in Hazelwood: Statistics, employment discrimination, and the 80 percent rule. *American Bar Foundation Research Journal,* 1, 139-164.

National Computer Systems, The Psychological Corporation and Measurement Incorporated. (1994, October 28). *Response to paper entitled: TAAS and accountability: Review, analysis, & policy implications.* Presented to the 8th annual Collaborative Research Conference. The University of Houston.

Plake, B.S. (1995). Differential item functioning in licensure tests. In J.C. Impara (Ed.). *Licensure testing: Purposes, procedures, and practices,* (pp. 205-218). Lincoln, Neb.: Buros-Nebraska series on Measurement and Assessment.

Shapiro, M.M. (1998, November 23). *Declaration of Martin M. Shapiro.*

Shepard, L.A. (1987, January). The Golden Rule agreement: Bad law, bad science. *The Score,* pp. 7,9.

Texas Student Assessment Program *Technical Digest* for the Academic Year 1996-1997.

Twing, J.S. (1998, December 13). *Cut Score SEMs.* Memo to S.E. Phillips.

Declaration of Rosalie Pedalino Porter

I, Rosalie Pedalino Porter, do state and declare as follows:

1. My name is Rosalie Pedalino Porter. I am the editor of a scholarly journal on the education of language minority students, *READ Perspectives,* published by the Institute for Research in English Acquisition and Development (READ) in Washington, D.C.

2. I hold a doctorate in bilingual education and English as a Second Language (ESL) conferred by the University of Massachusetts (Amherst) in 1982, as well as a master's of education (1979) and bachelor of arts, *magna cum laude* (1974) from the same institution. From 1979-80 I was a visiting scholar at the University of London Institute of Education, Department of English as a Foreign Language.

3. I am at present a member, by appointment of the governor, of the Massachusetts Education Reform Review Commission. I am also a member of two Massachusetts commissions concerned with statewide curriculum, instruction and assessment for language minority students: the English Language Learners Focus Group, and the State Advisory Council on Bilingual Education. From 1985-88, I served, by appointment of the United States Secretary of Education, as a member of the National Advisory and Coordinating Council on Bilingual Education, of the United States Department of Education. I was a research fellow at the Mary Ingraham Bunting Institute at Radcliffe College in 1987-88 and a Fulbright lecturer under the auspices of the United States Department of State in Rome, Italy, in 1992-93. I am a former member of the Executive Board of the Massachusetts Association of Teachers of English to Speakers of Other Languages (MATSOL), 1987-89, and former Chair of the Program Administrators' Group of the International Association for Teachers of English to Speakers of Other Languages (TESOL), 1986-87. I am the author of *Forked Tongue: The Politics of Bilingual Education* (Basic Books, 1990; 2nd. edition, Transaction Publishers, 1996). I have authored numerous articles on the topic of educating English language learners. A copy of my curriculum vitae is appended to this Declaration as Attachment A.

4. My professional experience includes five years as a Spanish bilingual and English as a Second Language (ESL) teacher in the Springfield, Mass., Public Schools from 1974-79. From 1980-90, I was the coordinator of bilingual and ESL programs for the Newton Public Schools in Newton, Mass., for children in nursery school through 12th grade. From 1993-98, in addition to my work as director of research for the READ Institute, I have

been a consultant to a number of school districts seeking to develop, evaluate and improve their programs for Limited-English Proficient (LEP) students, including districts in the states of California, Florida, Massachusetts, New York, Pennsylvania, Texas and Washington. I am familiar with the scholarly research in the field of bilingual education and educational services for Limited-English Proficient students, and am a frequent speaker and writer on the challenges in educating language minority students, principally children of Spanish-speaking background.

5. Since 1994 I have focused my professional work more closely on the collection and analysis of data on the academic progress of LEP students, both in English language learning and subject matter learning, and on what constitute fair and equitable guidelines for charting student achievement. The decision in *Castaneda v. Pickard* (648 F.2nd 989, Fifth Circuit, 1981) set three necessary conditions for school district compliance with the Equal Educational Opportunities Act: (1) instructional programs must follow an accepted educational theory; (2) adequate resources must be provided to implement the theory; and (3) the effectiveness of the program must be demonstrated by evidence of student academic progress in a reasonable amount of time. The third Castaneda condition—the accountability element—is generally conceded to be the crucial element of bilingual program evaluation that is often lacking.

6. Two examples of the lack of accountability in this area follow. Both California, the state enrolling 43 percent of all LEP students in the U.S., and Massachusetts, the state that first enacted a bilingual education law in 1971, have published reports documenting the lack of student assessment or data collection by their state education departments (*Meeting the Challenge of Language Diversity: An Evaluation of Programs for Pupils with Limited Proficiency in English,* California, 1992, and *Striving for Success: The Education of Bilingual Pupils, A Report of the Massachusetts Bilingual Education Commission,* 1994). California instituted a statewide testing program in 1998 that requires all students, including LEP children who have been in California one year or longer, to participate. In Massachusetts, meeting the legal obligation to collect and report data on LEP student achievement is now a state objective under the 1993 Education Reform Act.

7. The Massachusetts Education Reform Act in 1993 initiated the Massachusetts Comprehensive Assessment System (MCAS) to develop curriculum frameworks for all grade levels and in all school subjects, and to evaluate achievement of all students at the fourth-, eighth-, and 10th-grade levels (with an annual third grade reading test, the Iowa Test of Basic Skills, starting in 1997). The 10th-grade assessment is a "high stakes" test requiring a passing grade for high school graduation, beginning in 2002. In 1995, I helped write the guidelines for the participation and assessment of LEP students (when and under what conditions) in the English Language Arts

Frameworks, K-12. Massachusetts policy requires that LEP students who have been in U.S. schools three years or longer participate in the English Language Arts assessment.

8. I am a working member of the English Language Learners Focus Group whose ongoing mission is to advise the Massachusetts Department of Education on fair and equitable guidelines for the participation of LEP students in the MCAS assessments. The group recommendation that Spanish-language assessments in mathematics, science and technology, for those fourth-, eighth-, and 10th-grade LEP students who have been in U.S. schools fewer than three years has been adopted—all others are expected to take the tests in English. It is the expectation of Massachusetts educators that each year the MCAS tests are administered more data will be gathered on the areas at each grade level and in each subject where improvements are needed. This information will help to determine what additional resources are needed in which districts to improve student performance, i.e., staff development, curricular modifications, technology upgrades, scheduling more or less time in certain subjects.

9. During my 10 years as coordinator of programs for LEP students in the Newton, Mass., Public Schools, I supervised programs in two high schools for limited-English students from two dozen or more language backgrounds, most arriving in the U.S. with little or no knowledge of English. The Newton schools provided intensive English language courses and modified subject matter instruction, with the goal of helping these students meet the standards for high school graduation. Over 90 percent of the students who entered at high school age from other countries completed high school in two, three, or four years. Neither course requirements nor learning expectations were lowered for limited-English students.

10. Exempting whole groups of students from statewide assessments on the expectation that they will not perform adequately is unfair to the students who are excluded, as well as to their classmates. It has been my experience of 15 years as a bilingual teacher and program administrator that the majority of English language learners want to be included in the same education and testing programs as native English speakers and that they feel demeaned when they are excluded. A policy of separating language minority students, many of whom are native born, from the rest of the student population when the TAAS is administered is more likely to stigmatize and negatively impact the self-esteem of these students than is their inclusion in the tests. A past history of discrimination against Mexican-American and African-American children is not justification for holding these students to lower standards. According to Dr. Jose Cardenas, Texas has done much to eliminate discriminatory practices in the education of minority students in the past two decades. Maintaining rigorous standards and high expectations for minority students requires that periodic assessments of

each student's progress be conducted and reported. The useful data collected annually is used not only to improve teaching and learning but also to modify the testing program itself, as is the case with TAAS.

11. In my professional opinion, the Texas Education Agency's development and implementation of the Texas Assessment of Academic Skills (TAAS) plays an important part in meeting the Castaneda standard for evaluating academic progress of students who enter Texas public schools with a limited knowledge of the English language. I base my opinion on a study of Texas documents (listed in # 15 below), on my professional experience with school districts across the country, on my reading of the literature on accountability, standards and curricular improvement, and on my current involvement in Massachusetts efforts to consistently and systematically record, analyze, and chart the academic progress of LEP students.

12. The TAAS program, in my opinion, is a fair test of student learning, and the extensive reporting of student performance by subjects, grade levels, districts, and special populations provides a comprehensive, detailed array of information essential for a flexible, responsive educational system. The exit test administered to 10th-graders is, in my opinion, a reasonable assessment of essential skills that all high school graduates should have mastered, at a minimum. It is a reasonable test for those students who began their schooling in Texas as limited-English speakers. The provision of multiple opportunities to retake the test with remedial help is fair indeed. I believe it is sound educational policy to require one objective, uniform measure of student achievement as a prerequisite for high school graduation, an assessment closely based on the material taught in the schools.

13. As reported in the Texas Education Agency report of 1996-1997, minority students have registered consistently higher passing levels on the 10th-grade test each year since 1995, showing more rapid rates of improvement than for White non-Hispanic students. Disrupting the process of accountability for English-language learners would be a disservice to a group of students whose academic progress had not been monitored heretofore in a consistent, longitudinal, manner. To suggest that students should be granted high school diplomas without demonstrating minimal knowledge and skills on a uniform measure is not acceptable for the current requirements of the technological/information age job market or for pursuing higher education. Delia Pompa, director of the Office of Bilingual Education and Minority Languages Affairs in the U.S. Department of Education, commented pointedly on the need for LEP students to be held to reasonable learning standards and assessments: "I'm not sure it's O.K. for our kids to dance out something where other kids have to write on a subject to show mastery" (*Education Week*, May 18, 1994).

14. The complaint of plaintiffs' expert Linda McNeil that teaching time is devoted to "teaching the test" and not to a variety of more creative

instruction appears to me to be a harmful exaggeration. It is essential that students be taught "test-taking skills" in order to compete fairly with classmates who may have had more experience with standardized tests, but these skills need only be taught once and not year after year. As a former teacher I can state with confidence that a certain amount of review and sampling of test items is productive and is part of the learning process—not a waste of time. No competent teacher will spend all her time on test preparation to the exclusion of presenting the necessary subjects and arts and physical education activities.

15. I have reviewed the following documents: the *First Amended Complaint in the United States District Court for the Western District of Texas, San Antonio Division, G. I. Forum et al. v. Texas Education Agency, et al.*; the National Computer Systems *Annual Report 1991-92,* as prepared by the Austin Operations Center; the *Texas Student Assessment Program: Student Performance Results 1996-1997,* Texas Education Agency, Austin, Texas; sample TAAS Exit Level test administered March 1995; plaintiffs' experts reports by J. Cardenas, W. Haney, L. McNeil, R. Valencia and A. Valenzuela; and *Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995-1998,* November 1998.

16. I have served as an expert witness in a number of court cases in the area of the education of limited-English students (see Appendix A, page 5). Within the past four years I have been deposed in *Sang Van et al. v. Seattle School District* (1994-95) for the defendants, and in *Carbajal et al. v. Albuquerque Public School District* (1998) for the plaintiffs.

References

BW Associates. (1992). *Meeting the Challenge of Language Diversity: An Evaluation of Programs for Students with Limited Proficiency in English.* Berkeley, Calif.

Dev/Tech Systems, Inc. (1996). *A Descriptive Study of the ESEA Title VII Educational Services Provided for Secondary School Limited-English Proficient Students.* Washington, D.C.: Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education.

Glenn, Charles L. (1997). "What Does the National Research Council Study Tell Us about Educating Language Minority Children?" *READ Perspectives,* Vol. IV-2 Fall 1997, pp. 66-95.

Hollander, Paul. (1998). "New Insights: A Review of Ruben Rumbaut's 'Transformations: The Post Immigrant Generation in an Age of Diversity,'" *READ Perspectives,* Vol. V-2, Fall 1998, pp. 55-64.

Lopez, Mark H. and Mora, Marie T. (1998). "The Labor Market Effects of Bilingual Education Among Hispanic Workers," *READ Perspectives,* Vol. V-2, Fall 1998, pp. 33-54.

Massachusetts Bilingual Education Commission. (1994). *Striving for Success: The Education of Bilingual Pupils.* Boston: Commonwealth of Massachusetts.

Porter, Rosalie P. (1995) *Forked Tongue: The Politics of Bilingual Education,* 2nd. Edition. New Brunswick, N.J.: Transaction Publishers, Rutgers University.

__________. (1995). "A Review of the U.S. General Accounting Office Study of Limited-English Proficient Students." *READ Perspectives,* Vol. II-1, Spring 1995, pp. 9-23.

__________. (1994). "Goals 2000 and the Bilingual Student," *Education Week,* May 18, 1994, pp. 45 36.

Rossell, Christine M. and Baker, K. (1996) *Bilingual Education in Massachusetts: The Emperor Has No Clothes.* Boston: The Pioneer Institute.

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF TEXAS
SAN ANTONIO DIVISION

GI FORUM, IMAGE DE TEJAS, §
RHONDA BOOZER, MELISSA §
MARIE CRUZ, MICHELLE §
MARIE CRUZ, LETICIA ANN §
FAZ, ELIZABETH GARZA, §
MARK GARZA, ALFRED LEE §
HICKS, BRANDYE R. JOHNSON, §
JOCQULYN RUSSELL, §
§
Plaintiffs, §
§
vs. § Civil Action No. SA-97-CA-1278-EP
§
TEXAS EDUCATION AGENCY, §
DR. MIKE MOSES, MEMBERS, §
AND THE TEXAS STATE §
BOARD OF EDUCATION, §
in their official capacities, §
§
Defendants, §

JUDGMENT

In accordance with this Court's opinion of this same date, it is hereby ORDERED, ADJUDGED, and DECREED that judgment is entered in favor of the Defendants and against the Plaintiffs. All costs are to be borne by the parties incurring them. It is further ORDERED that all pending motions be stricken from the docket as moot and that this case is DISMISSED.

SIGNED and ENTERED this 7th day of January 2000.
[signature]
EDWARD C. PRADO
UNITED STATES DISTRICT JUDGE

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF TEXAS
SAN ANTONIO DIVISION

GI FORUM, IMAGE DE TEJAS,)
RHONDA BOOZER, MELISSA)
MARIE CRUZ, MICHELLE)
MARIE CRUZ, LETICIA ANN)
FAZ, ELIZABETH GARZA,)
MARK GARZA, ALFRED LEE)
HICKS, BRANDYE R. JOHNSON,)
JOCQULYN RUSSELL,)
)
Plaintiffs,) Civil Action SA-97-CA-1278-EP
)
vs.)
)
TEXAS EDUCATION AGENCY,)
DR. MIKE MOSES, MEMBERS,)
AND THE TEXAS STATE)
BOARD OF EDUCATION,)
in their official capacities,)
Defendants.)

ORDER

The issue before the Court is whether the use of the Texas Assessment of Academic Skills (TAAS) examination as a requirement for high school graduation unfairly discriminates against Texas minority students or violates their right to due process. The Plaintiffs challenge the use of the TAAS test under the Due Process Clause of the United States Constitution and 34 C.F.R. § 100.3, an implementing regulation to the Title VI of the Civil Rights Act of 1964, asking this Court to issue an injunction preventing the Texas Education Agency (TEA) from using failure of the exit-level TAAS test as a basis for denying high school diplomas.[1] The Court has considered the testimony and evidence presented during five weeks of trial before the bench, as well as the relevant case law. After such consideration, and much reflection, the Court has determined that the use of the TAAS examination does not have an impermissible adverse impact on Texas's minority students and does not violate their right to the due process of law. The bases for the Court's determination are outlined more fully in its findings of facts and conclusions of law, below. The Court writes separately only to make a few general observations about the legal issues underpinning this case.

In deciding the issues presented, both at the summary judgment stage and at trial, the Court has been required to apply a body of law that has not always provided clear guidance. It is clear that the law requires courts to give deference to state legislative policy, *see Board of Educ. v. Mergens,* 496 U.S. 226, 251 (1990); in the educational context, such deference is even more warranted, *see San Antonio Indep. Sch. Dist. v. Rodriguez,* 411 U.S. 1, 42 (1973). Education is the particular responsibility of state governments. *Id.* Moreover, courts do not have the expertise, or the mandate of the electorate, that would justify unwarranted intrusion in curricular decisions. *See id.* On the other hand, these considerations cannot be used to tie a court's hands when a state uses its considerable power impermissibly to disadvantage minority students.

This case requires the application of law from a number of diverse areas —employment law, desegregation law, and testing law in areas such as bar examinations or teacher certification examinations. Only one case cited by any party or this Court is both controlling and directly on point—*Debra P. v. Turlington,* 644 F.2d 397 (5th Cir. 1981). In *Debra P.,* the United States Court of Appeals for the Fifth Circuit found that a state could overstep its bounds in implementing standardized tests as graduation requirements. Specifically, the court found that a test that did not measure what students were actually learning could be fundamentally unfair. The court also found that a test that perpetuated the effects of prior discrimination was unconstitutional. This Court finds these ideas to be in step with the United States Supreme Court's suggestion in *Regents of University of Michigan v. Ewing,* 474 U.S. 214, 225 (1985), that a state could violate the Constitution if it implemented policies that violated accepted educational norms.

In addition, this Court has allowed the Plaintiffs to bring a claim pursuant to a regulation adopted in conjunction with Title VI. *See* 34 C.F.R. § 100.3. That regulation, in clear, unmistakable terms, prohibits a federally funded program from implementing policies that have a disparate impact on minorities. *Id.* While the Court acknowledges that the United States Supreme Court has limited Title VI itself to constitutional parameters (i.e., has required a showing of an intent to discriminate in order to prove a violation), see *United States v. Fordice,* 505 U.S. 717, 722 n.7 (1992), the Court does not find that this limitation has been clearly and unambiguously extended to its implementing regulations. The Court is not alone in reaching this conclusion. *See Cureton v. National Collegiate Athletic Assoc.,* No. 99-1222, 1999 WL1241077, at *5 (3d Cir. Dec. 22, 1999); *Elston v. Talladega Co. Bd. of Educ.,* 997 F.2d 1394, 1406 (11th Cir. 1993); *Harper v. Board of Regents of Ill. State Univ.,* 35 F. Supp.2d 1118, 1123 (C.D. Ill. 1999); *Valeria G. v. Wilson,* 12 F. Supp.2d 1007, 1023 (N.D. Cal. 1998); *Graham v. Tennessee Secondary Athletic Ass'n,* No. 1:05-CV-044, 1995 WL 115890, at *12 (E.D. Tenn. Feb. 20, 1995). Nor is the court alone in concluding that a private right of action exists under this regulation. See, e.g., *Harper*, 35 F. Supp.2d at 1123; *Valeria G.*, 12 F, Supp.2d at 1023; *Graham*, No. 1:05-CV-044, 1995 WL 115890, at *12. The Court believes that it has followed the law as it presently exists in allowing these claims to go forward.

In reviewing the diverse cases that underpin this decision, the Court had to acknowledge what the Defendants have argued throughout trial—this case is, in some important ways, different from those cases relied upon by the Plaintiffs. In the first place, this case asks the Court to consider a standardized test that measures knowledge rather than one that predicts performance. The Court has had to consider whether guidelines established in the employment context are adequate for determining whether an adverse impact exists in this context. In addition, the Court has been required to determine the deference to be given to a State in deciding *how much* a student should be required to learn—the cut-score issue. Finally, the Court has had to weigh what appears to be a significant discrepancy in pass scores on the TAAS test with the overwhelming evidence that the discrepancy is rapidly improving and that the lot of Texas's minority students, at least as demonstrated by academic achievement, while far from perfect, is better than that of minority students in other parts of the country and appears to be getting better.[2]

This case is also remarkable for what it does *not* present for the Court's consideration. In spite of the diverse and contentious opinions surrounding the use of the TAAS test, this Court has not been asked to—and indeed could not—rule on the wisdom of standardized examinations. This Court has no authority to tell the State of Texas what a well-educated high school graduate should demonstrably know at the end of twelve years of education. Nor may this Court determine the relative merits of teacher evaluation and "objective" testing.

This case is also not directly about the history of minority education in the State. While that history has had some bearing on some of the due process concerns raised by the Plaintiffs, what is really at issue here is whether the TAAS exit-level test is *fair*. As the Court notes below, the test cannot be fair if it is used to punish minorities who have been victimized by state-funded unequal educations. Thus, the Court has carefully considered the claims that Texas schools still offer widely diverse educational opportunities and that, too often, those opportunities depend on the color of a student's skin or the financial resources of the student's school district.[3] To some degree, as discussed below, the Court must accept these claims. But that finding, alone, is an insufficient basis for invalidating this examination. There must be some link between the TAAS test and these disparities. In other words, the Plaintiffs were required to prove, by a preponderance of the evidence, that the TAAS test was implemented in spite of the disparities or that the TAAS test has perpetuated the disparities, and that requiring passage of the test for graduation is therefore fundamentally unfair. The Court believes that this has not been proven. Instead, the evidence suggests that the State of Texas was aware of probable disparities and that it designed the TAAS accountability system to reflect an insistence on standards and educational policies that are uniform from school to school. It is true that these standards reflect no more than what the State of Texas has determined are essential skills and knowledge. It is undeniable that there is more to be learned. However, the Court cannot pass on the State's determination of what, or how much, knowledge must be acquired prior to high school graduation.

This case presented widely differing views of how an educational system should work. One set of witnesses believed that the integrity of objective measurement was paramount; the other believed that this consideration should be tempered with more flexible notions of fairness and justice. Thus, the relative quality of experts in this case is not so simple a matter as either party would make it. On the issue of internal test fairness and soundness, clearly the TEA presented better experts—their experts wrote the test and have written other tests. Their experts are invested in the profession and practice of test-writing and are committed to standardized tests as useful exercises for various kinds of educational measurement. However, TEA's experts were not so qualified, the Court finds, to speak on the wisdom of the use of standardized tests as they apply to ethnic minorities in a state educational system that has had its difficulties providing an equal education to those minorities. In that regard, the expert testimony failed to match up. TEA's experts, for example, are not especially qualified to speak on the psychological, social, or economic effects of failing to pass a test used as a requirement for graduation. At least one of those experts testified that whether a given test item disadvantages minority students is a factor that an item reviewer may ultimately *reject* in determining whether an otherwise valid item should be placed on the test. This is so because, as TEA's experts overwhelmingly testified, what is fundamentally important to these psychometricians is that the test objectively measure the material that it purports to measure and that it measure content that students have been exposed to.[4] *See Report of Dr. Susan Phillips,* Defendants' expert, at 16 (a plausible explanation for differential performance is the difference in achievement level). On the question, then, of whether it is wise to use standardized tests in making high-stakes decisions, taking into account all the contextual factors, the Court finds the expert testimony was not fairly joined. Plaintiff's experts had clearly considered this question more fully and given it more weight. The question is—how relevant to this Court's decision is the *wisdom* of the TAAS test and, to the extent that Plaintiff's experts were able to prove that the test is not *wise,* have they been able to show that it actually crosses the line and is impermissible by some legal standard?

Ultimately, resolution of this cases turns not on the relative validity of the parties' views on education but on the State's right to pursue educational policies that it legitimately believes are in the best interests of Texas students. The Plaintiffs were able to show that the policies are debated and debatable among learned people. The Plaintiffs demonstrated that the policies have had an initial and substantial adverse impact on minority students. The Plaintiffs demonstrated that the policies are not perfect. However, the Plaintiffs failed to prove that the policies are unconstitutional, that the adverse impact is avoidable or more significant than the concomitant *positive* impact, or that other approaches would meet the State's articulated legitimate goals. In the absence of such proof, the State must be allowed to design an educational system that it believes best meets the need of its citizens.

FINDINGS OF FACT AND CONCLUSIONS OF LAW

FINDINGS OF FACT [5]

THE TEST

Test Construction

In 1984, the Texas legislature passed the Equal Educational Opportunity Act (EEOA), designed to impose an "accountability" system on Texas public school administrators, teachers, and students. The following year, in response to that legislation, the Texas State Board of Education adopted a curriculum of Essential Elements.[6] In addition, the Board moved forward with its plans to implement an objective standardized test that would measure mastery of the state-mandated curriculum. In 1987, Texas instituted the TEAMS high school graduation exit test, given to eleventh-graders.

In 1990, Texas replaced the TEAMS test with the Texas Assessment of Academic Skills (TAAS) test, the subject of this lawsuit. Like the TEAMS test, the TAAS test is designed to measure mastery of the state-mandated curriculum. However, the TAAS test seeks to assess higher-order thinking and higher problem-solving skills than did the TEAMS test. The TAAS test is developed and constructed by National Computer Systems (NCS), a private corporation. NCS, in turn, subcontracts development of TAAS items to Harcourt Brace Educational Measurement (HBEM) and Measurement Incorporated. HBEM contracts with individuals to write items for the TAAS test. In addition to the extensive input from these professional test-designers, many of whom are not in the State of Texas, there is a great deal of input from state educators in the design of the TAAS test. Decisions as to which portions of the state-mandated curriculum should be measured by the TAAS test are made by Texas teachers and educational professionals. The Texas Education Agency has ensured that the educators comprise an ethnically diverse group of individuals from across the state. In addition, proposed TAAS questions are reviewed by subject-matter content experts, review committees of teachers and educators, test-construction experts, and measurement experts.

In reviewing test items, educators are instructed to consider the following issues: relevancy of the item, difficulty range, clarity of the item, correctness of the keyed answer choice, and the plausability of distractors. Reviewers are also asked to consider the more global issues of passage appropriateness, passage difficulty, and interactions between items within and between passages as well as work, graphs, or figures. Reviewers are asked to assess whether or not each item on the TAAS exam covers information that was sufficiently taught in the classroom by the time of the test administration. After the initial review, a second review is conducted by staff members of the Student Assessment and Curriculum Divisions of the TEA and by developmental and scoring contractors.

Selected questions are then field tested. The results of those field tests are reviewed by a Data Review Committee. Committee members are permitted to remove items they consider to be questionable, including questions that a disproportionate number of minority students fail to answer correctly. Reviewing members are given "great deference" in this process and are not required to eliminate a question that reflects that any ethnic group had particular difficulty with the question. *See Report of Dr. Susan Phillips,* Defendants' expert, at 17. If the reviewer finds that an item with a predicted adverse effect on minorities is a "fair measure of its corresponding state objectives *for all students,* and is free of offensive language or concepts that may differentially disadvantage minority students," the item may be retained, even if a significantly large number of minority students do no answer it correctly. *Id.* (emphasis in original).

Test Validity

Several concepts are key to understanding the arguments raised by the parties regarding the validity of the TAAS examination. The "validity" of a given standardized test refers to the "weight of the accumulated evidence supporting the particular use of the test scores." *Report of Dr. Susan Phillips,* Defendants' expert, at 3. "Content validity" measures the degree to which the test measures the knowledge and skills sought to be measured, in this case the legislatively mandated minimum essentials. *Id.* "Curricular validity" refers to the issue of whether students have an adequate opportunity to learn the material covered on a given standardized test. *Id.* at 10. "Test reliability" is "an indicator of the consistency of measurement." *Id.* at 4. Reliability may be tested by repeat testing or by various measures based on a single-test measurement. *Id.*

Each form of a standardized test must be valid and reliable. Validity and reliability across different forms of the test are ensured by "equating" test forms, or adjusting for any minor variations in difficulty between the forms. *Id.* at 7. The TAAS test is "equated" under what is called the Rasch Model. *Id.* This model focuses narrowly on item-difficulty parameters and does not provide for "item weighing," as do more complex equating models. *Id.* In other words, part of equating test forms involves using a fairly simple formula, the Rasch Model, to determine how well a student's response on a given question predicts that student's success on the exam as a whole. "Point biserials" measure the degree to which persons who answer an item correctly tend to also have high total test scores and vice versa. *Id* at 21.

Test Administration

Texas public school students begin taking the TAAS test in the third grade. In the tenth grade, Texas public school students are given what is called the "exit-level" TAAS exam, or the examination they must pass in order to graduate. Students must pass each of three portions of the TAAS test—a reading, mathematics, and writing portion—in order to graduate. Texas public school students who do not pass the test on their first attempt are then given at least seven additional opportunities to take and pass the TAAS exam before their scheduled graduation date.

THE PASSING STANDARD

The initial passing standard, or cut score, on the TAAS test was set at 60 percent, and a 70-percent passing standard was phased in after the first year. In setting the passing standard, the State Board of Education looked at the passing standard for the TEAMS test, which was also 70 percent, and also considered input from educator committees. In addition, the selection of the score reflected a general sense that 70 percent of the required essential elements was sufficient "mastery" for the purposes of graduation. *See TEA Board of Education Minutes,* June 1990.

The TEA understood the consequences of setting the cut score at 70 percent. When it implemented the TAAS test, the TEA projected that, with a 70-percent cut score, at least 73 percent of African Americans and 67 percent of Hispanics would fail the math portion of the test; at least 55 percent of African Americans and 54 percent of Hispanics would fail the reading section; and at least 62 percent of African Americans and 45 percent of Hispanics would fail the writing section. The predictions for white students were 50 percent, 29 percent, and 36 percent, respectively. However, TEA representatives had reason to believe that those projections were inflated. Experts informed TEA representatives that there is a measurable difference in the motivation between students taking a field examination and students taking a test with actual consequences. While the passing numbers were somewhat better than projected, they were nonetheless alarming. On the October 1991 administration of the exam to tenth graders, 67 percent of African Americans and 59 percent of Hispanics failed to meet the passing cut score. For whites, the number was 31 percent.

OBJECTIVE MEASUREMENT

In spite of projected disparities in passing rates, the TEA determined that objective measures of mastery should be imposed in order to eliminate what it perceived to be inconsistent and possibly subjective teacher evaluations of students. The TEA offered evidence at trial that such inconsistency exists. The TEA also presented testimony that subjectivity can work to disadvantage minority students by allowing inflated grades to mask gaps in learning.

REMEDIATION

Failure to master any portion of the exam results in state-mandated remediation in the specific subject area where the student encountered difficulty. There is no state-mandated approach to remediation, however. Consequently, remedial efforts vary from district to district. The evidence at trial reflected varying degrees of success resulting from remedial efforts. The Court finds that, on balance, remedial efforts are largely successful. TEA's expert, Dr. Susan Phillips, estimates that 44,515 minority students in 1997 were successfully remediated after failing their first attempt at the TAAS test in 1995. *Report of Dr. Susan Phillips,* Defendants' expert, at 14. The Court finds this evidence credible.

ACCOUNTABILITY

Administrators, schools, and teachers are held accountable, in varying degrees, for TAAS performance. The accountability system does not ignore the presence of ethnic minorities in the system or the difficulties minorities may have in passing the examination. Passing and failing scores are disaggregated, or broken down into subgroups, so that schools and districts are aware of the degree of success or failure of African American, Hispanic, and white students. If one subgroup fails to meet minimum performance standards, a school or district will receive a low accountability rating.

HISTORY OF TESTING/DISCRIMINATION IN TEXAS

It is beyond dispute that standardized tests have been used in educational contexts to disadvantage minorities. *See Report of Dr. Uri Treisman,* Defendants' expert, at 3. However, the Plaintiffs have presented insufficient evidence to support a finding that the TAAS test, as developed, implemented, and used in Texas, is designed to or does impermissibly disadvantage minorities. While it is true that a number of minority students fail to pass the TAAS test and earn a diploma, there is no evidence that this was the design of the State in initiating the test. On the contrary, there is evidence that one of the goals of the test is to help identify and eradicate educational disparities. The receipt of an education that does not meet some minimal standards is an adverse impact just as surely as failure to receive a diploma.

The Court agrees with Plaintiffs that sufficient evidence, including evidence cited in other state and federal case law, exists to support the Plaintiffs' claim that Texas minority students have been, and to some extent continue to be, the victims of educational inequality. *See Report of Dr. Uri Triesman,* Defendants' Expert, at 7.; *see also e.g., United States v. Texas Educ. Agency,* 467 F.2d 848 (5th Cir. 1972), and its progeny; *United States v. Texas,* 330 F. Supp. 235 (E.D. Tex. 1971). Witnesses in this case were questioned by counsel and by the Court about the reasons for this inequality. The evidence was disturbing, but inconclusive. Socio-economics, family support, unequal funding, quality of teaching and educational materials, individual effort, and the residual effects of prior discriminatory practices were all implicated. The Court finds that each of these factors, to some degree, is to be blamed.

However, the Plaintiffs presented insufficient evidence to support a finding that minority students do not have a reasonable opportunity to learn the material covered on the TAAS examination, whether because of unequal education in the past or the current residual effects of an unequal system. The Plaintiffs presented evidence to show that, in a more general sense, minorities are not provided equal educational opportunities. In particular, Plaintiffs demonstrated that minorities are underrepresented in advanced placement courses and in gifted-and-talented programs. Minority students are also disproportionately taught by non-certified teachers. However, because of the rigid, state-mandated correlation between the Texas Essentials of Knowledge and Skills and the TAAS test, the Court finds that all Texas students have an equal opportunity to learn the items presented on the TAAS test, which is the issue before the Court. In fact, the evidence showed that the immediate effect of poor performance on the TAAS examination is more concentrated, targeted educational opportunities, in the form of remediation. Moreover, the TEA's evidence that the implementation of the TAAS test, together with school accountability and mandated remedial follow-up, helps address the effects of any prior discrimination and remaining inequities in the system is both credible and persuasive.

EDUCATIONAL STANDARDS

Current prevailing standards for the proper use of educational testing recommend that high-stakes decisions, such as whether or not to promote or graduate a student, should not be made on the basis of a single test score.

See Supplemental Report of Dr. Walter Haney, Plaintiff's expert, at 42 (citing *Standards for Educational and Psychological Testing* (1985). There was little dispute at trial over whether this standard exists and applies to the TAAS exit-level examination. What was disputed was whether the TAAS test is actually the sole criterion for graduation. As the TEA points out, in addition to passing the TAAS test, Texas students must also pass each required course by 70 percent. See TEXAS ADMIN. CODE § 74.26(c). Graduation, in Texas, in fact, hinges on *three separate and independent* criteria: the two objective criteria of attendance and success on the TAAS examination, and the arguably objective/subjective criterion of course success. However, as the Plaintiffs note, these factors are not weighed with and against each other; rather, failure to meet any single criterion results in failure to graduate. Thus, the failure to pass the exit-level exam does serve as a bar to graduation, and the exam is properly called a "high-stakes" test.

On the other hand, students are given at least eight opportunities to pass the examination prior to their scheduled graduation date. In this regard, a single TAAS score does *not* serve as the sole criterion for graduation. The TEA presented persuasive evidence that the number of testing opportunities severely limits the possibility of "false negative" results and actually increases the possibility of "false positives," a fact that arguably advantages all students whose scores hover near the borderline between passing and failing.

DISPARATE IMPACT

The Court finds an inescapable conclusion that in every administration of the TAAS test since October 1990, Hispanic and African American students have performed significantly worse on all three sections of the exit exam than majority students. However, the Court also finds that it is highly significant that minority students have continued to narrow the passing rate gap at a rapid rate. In addition, minority students have made gains on other measures of academic progress, such as the National Assessment of Educational Progress test. The number of minority students taking college entrance examinations has also increased.

In determining whether a legally significant statistical disparity exists, the Court has had to consider two difficult issues. The first is whether to apply the EEOC's Four-Fifths Rule or some other recognized test for identifying statistical disparity, as the Plaintiffs have argued the Court must do. The second is whether to consider cumulative pass rates or pass rates on a single administration of the examination at the tenth-grade level. The Court's resolution of these issues is discussed more fully in the Conclusions of Law, below.

Plaintiffs statistical expert, Mark Fassold, presented evidence that TAAS exit-level exam failure rates have a racially discriminatory effect under the

Four-Fifths Rule[7] and the *Shoben* formula.[8] The TEA contends that Fassold's study is flawed in significant ways and must be rejected. The Court acknowledges that Fassold's data include students who did not sit for the exam in the category of students who "passed" the exam. However, the Court has considered this flaw in its proper context. As the Plaintiffs point out, Fassold's methodology almost certainly artificially *inflates* the minority pass rate by coding those who fail to take the examination as passing. *Report of Mark Fassold,* Plaintiffs expert, at 13 n.10. Because minorities fail to take the test at a higher rate than majority students, the minority pass rate is inflated at a higher rate than that of the majority pass rate. *Id.* Thus, the Court is inclined to agree with Plaintiffs that they have likely *over*-estimated the minority pass rate. In this context, then, the Court finds there is sufficient evidence that, on first-time administration of the exit-level test, a legally significant adverse impact exists. While an examination of cumulative pass scores in more recent years does not evince adverse impact under the Four-Fifths Rule, the disparity there, too, is sufficient to give rise to legitimate concern. *See Cureton v. National Collegiate Athletic Assoc.*, 37 F. Supp.2d, 687, 697 (E.D. Pa. 1999) ("no rigid mathematical threshold of disproportionality... must be met to demonstrate a sufficiently adverse impact"), rev'd on other grounds, No. 99-1222, 1999 WL 1241077 (3d Cir. Dec. 22, 1999). Moreover, as discussed below, there are significant statistical disparities in cumulative pass rates.

In addition to evaluating the statistical impact of the examination, the Court has, at the behest of both parties, considered the "practical consequences" or "practical impact" of the high failure rates of minorities. That consideration involves careful examination of the immediate and long-term effects of the statistically disparate failure rates. The TEA argues that, because of the presence of largely successful remediation, the practical significance benefits minorities. The Plaintiffs note that failure to graduate has serious economic, social, and emotional effects on students.

The Court finds that failure of the exit-level TAAS examination during the first seven administrations results in immediate remedial efforts. At the last administration, of course, failure of the exit-level TAAS examination results in a failure to receive a diploma. However, the Court finds, based on the evidence presented at trial, that the effect of remediation, which is usually eventual success in passing the examination and thus receipt of a high school diploma, is more profound than the steadily decreasing minority/failure rate.

DROP-OUT/RETENTION RATES

Plaintiffs presented sufficient evidence to support a finding that Texas students, particularly minority students, drop out of school in significant numbers and are retained at their current grade level in numbers that give cause for concern. Moreover, the Plaintiffs presented evidence supporting their contention that drop-out and retention rates for minorities are peculiarly high at the ninth grade, just before the first administration of the exit-level TAAS. *See Supplemental Report of Dr. Walter Haney*, Plaintiff's expert, at 21-29. The evidence presented by Plaintiffs also shows that in the year 1991, as the present TAAS test was being phased in, there was a drop in the ratio of high school graduates to grade nine students three years before, and that this drop was most notable for minority students. See *Id.* at 25-26. However, Plaintiffs have failed to make a causal connection between the implementation of the TAAS test and these phenomena, beyond mere conjecture. In other words, Plaintiffs were only able to point to the problem and ask the Court to draw an inference that the problem exists because of the implementation of the TAAS test. That inference is not, in light of the evidence, inevitable. The Defendants hypothesize, just as plausibly, for example, that the ninth grade increase in drop outs is due to the cessation of automatic grade promotion at the beginning of high school in Texas.

CONCLUSIONS OF LAW[9]

This lawsuit is properly brought under two causes of action: the implementing regulations of Title VI of the Civil Rights Act of 1964 and the Due Process Clause of the Fourteenth Amendment to the United States Constitution.

TITLE VI REGULATIONS

Title VI of the Civil Rights Act of 1964 is a statute enacted "with the intent" to invoke the Fourteenth Amendment's congressional enforcement power." *Lesage v. State of Texas*, 158 F.3d 213, 218 (5th Cir. 1998), *cert. filed*, 67 USLW 3469 (Jan. 11, 1999). The TEA, as a state agency that administers and monitors compliance with educational programs required by state and federal laws and as the recipient of federal funds, is governed by Title VI and its regulations. 42 U.S.C. 2000d et seq.; *Castenada v. Pickard*, 648 F.2d 989, 992 (5th Cir. Unit A 1981). The Plaintiffs have brought this suit, in part, pursuant to 34 C.F.R. § 100.3, a regulation promulgated by the Department of Education to implement Title VI. That regulation prohibits activity in federally funded programs that has the effect of subjecting individuals to discrimination because of their race, color, or national origin. 34 C.F.R. § 100.3; *Powell v. Ridge*, 189 F.3d 387, 396 (3d Cir. 1999), *cert. denied*, 1999 WL 783927 (Dec. 6, 1999), *Elston*, 997 F.2d at 1406. The language of the regulation clearly suggests that a disparate impact analysis is appropriate under this regulation, and courts have applied it in that manner.[10] *See Quarles v. Oxford Mun. Separate Sch. Dist.*, 868 F.2d 750, 754 n.3 (5th Cir. 1989); *City of Chicago v. Lindley*, 66 F.3d 819, 827 (7th Cir. 1995); *see also Cureton*, 37 F. Supp. 2d at 697 (gathering cases). Similarly, courts have held that plaintiffs bringing lawsuits pursuant to 34 C.F.R. § 100.3 have a private right of action. *Powell,* 189 F.3d at 398; *Cureton,* 37 F. Supp.2d at 689. This Court concurs in that conclusion.

A disparate impact theory of racial discrimination permits a court to overturn facially neutral acts and policies that have "significant adverse effects on protected groups...without proof that the [actor] adopted those practices with a discriminatory intent." *Watson v. Fort Worth Bank and Trust,* 487 U.S. 977, 986-87 (1988). To delineate a standard for evaluating this disparate impact claim, the Court has looked to employment law under Title VII of the Civil Rights Act of 1964, which allows a disparate impact cause of action. *See, e.g., Wards Cove Packing Co., Inc. v. Atonio,* 490 U.S. 642 (1989); *Watson,* 487 U.S. 977; *Griggs v. Duke Power Co.,* 401 U.S. 424 (1971).

Thus, in determining whether a prima facie case of disparate impact has been established, this Court will apply the burden-shifting analysis established in Title VII cases. Under that analysis, the plaintiff must initially demonstrate that the application of a facially neutral practice has caused a disproportionate adverse effect. *Wards Cove,* 490 U.S. at 656-57. If a plaintiff makes such a showing, a burden of production shifts to the defendant. Under that burden, the defendant must produce evidence that the practice is justified by an educational necessity. *Id.* The plaintiff may then ultimately prevail by demonstrating that an equally effective alternative practice could result in less racial disproportionality while still serving the articulated need. *Watson,* 487 U.S. at 998.

I. Disparate Impact

In determining whether an adverse impact exists in this case, the Court has considered and applied the Equal Employment Opportunity Commission's Four-Fifths Rule. *See* 29 C.F.R. § 1607.4(d)/. The Court disagrees with the TEA's argument that this test is not suited for identifying the presence of adverse impact in this context. *See Cureton,* 37 F. Supp.2d at 700 (applying Four-Fifths Rule). In addition, the Court notes that the TEA did not offer in its briefing or at trial a satisfactory substitute for determining a statistical disparity, choosing instead to rely on its arguments that a disparate impact theory should not be applied in a Title VI case or, alternatively, that the Court should consider only the practical effect of remediation.

In addition to the Four-Fifths Rule, the Court has considered the statistical significance of the observed differences in pass rates. The methodology for such consideration, referred to by these parties as the *Shoben* formula, is to find a "z-score," or a number representing the differences between independent proportions—here the pass rates of minority students and the pass rates of majority students. *See Report of Mark Fassold,* Plaintiff's expert, at 4-6; *Preliminary Report of Dr. Walter Haney,* Plaintiff's expert, at 13.

The evidence regarding whether Plaintiffs have established the existence of a significant adverse impact on minority students is mixed. Plaintiffs' statistical analysis, while somewhat flawed, demonstrates a significant impact on first-time administration of the exam. This impact, which clearly satisfies the Four-Fifths Rule, is conceded by at least one TEA expert. *See Report of Dr. Susan Phillips,* Defendants' expert, at 13. However, cumulative pass rates do not demonstrate so severe an impact and, at least for the classes of 1996, 1997, and 1998, are not statistically significant under the EEOC's Four-Fifths Rule. *See Id.* at 14.

In considering how to handle the dilemma of choosing between cumulative and single-test administration, the Court has taken into account the immediate impact of initial and subsequent in-school failure of the exam—largely successful educational remediation. In addition, the Court has considered the evidence that minority scores have shown dramatic improvement. These facts would seem to support the TEA's position that cumulative pass rates are the relevant consideration here.

The Plaintiffs argue that successful remediation and pass-rate improvement should not be considered in determining whether an adverse impact exists. To support their argument, the Plaintiffs point to case law holding that a "bottom line" defense is insufficient to combat a showing of adverse impact. *See Connecticut v. Teal,* 457 U.S. 440, 455 (1982). The Court is not convinced that this argument is applicable to the case before it.

In *Connecticut v. Teal,* the United States Supreme Court held that an employer charged with a Title VII violation could not justify discrimination against one individual by pointing to its favorable treatment of other members of the same racial group. *Id.* at 454. According to the Court, Title VII requires an employer to provide "an equal opportunity for each applicant regardless of race." *Id.* In that case, however, the employer was trying to compensate for a discriminatory selection test by arguing that subsequent affirmative action practices allowed the employer to reach a non-discriminatory "bottom-line." *Id.* at 452-53. As another court has stated, Teal stands for the proposition that "the disparate exclusion of minority candidates at the first stage of the selection process was not ameliorated by the favorable end result because excluded candidates were deprived individually of the opportunity for promotion." *Lindley*, 66 F.3d at 829.

The Court will assume that *Teal's* analysis applies in Title VI cases. *Id.* However, the Court is not sure that *Teal* is relevant here. Failure to pass the first administration of the TAAS test does not deny an individual a competitive opportunity. It is only after at least *eight* tries that there is a real negative impact. This is not a case where there are several distinct steps through a selection system. *See Newark Branch, NAACP v. Town of Harrison, N.J.*, 940 F.2d 792, 801 (3d Cir. 1991). Nor is it the TEA's argument that the test is legal because, while some individuals fail and do not receive diplomas, others do and so the disparate effect is ameliorated. Rather, the TEA is arguing that each individual student is given at least eight tries to pass the exam and that many students who fail on the first attempt eventually succeed. The Court believes that these facts distinguish this case from *Teal,* and the Court will reject the *Teal* analysis. Thus, the Court has considered, and found relevant, the distinction between pass rates after a single administration and pass rates after eight attempts.

Having said all that, however, the Court finds that, whether one looks at cumulative or single-administration results, the disparity between minority and majority pass rates on the TAAS test must give pause to anyone looking at the numbers. The variances are not only large and disconcerting, they also apparently cut across such factors as socioeconomics. Further, the data presented by the Plaintiffs regarding the statistical significance of the disparities buttress the view that legally meaningful differences do exist between the pass rates of minority and majority students. Disparate impact is suspected if the statistical significance test yields a result, or z-score, of more than two or three standard deviations. *Castenada v. Partida,* 430 U.S. 482, 496 n.17 (1977). In all cases here, on single and cumulative administrations, there are significant statistical differences under this standard. Given the sobering differences in pass rates and their demonstrated statistical significance, the Court finds that the Plaintiffs have made a prima facie showing of significant adverse impact. *See Supplemental Report of Dr. Walter Haney,* Plaintiffs Expert, at 4-5 (discussing practical adverse impact); *Cureton,* 37 F. Supp.2d at 697 ("no rigid mathematical threshold of disproportionality...must be met to demonstrate a sufficiently adverse impact").

II. Educational Necessity

Having found that the Plaintiffs have established a prima facie showing of significant adverse impact, the Court must consider whether the TEA has met its burden of production on the question of whether the TAAS test is an educational "necessity." The word "necessity," as an initial matter, is somewhat misleading; the law does not place so stringent a burden on the defendant as that word's common usage might suggest. Instead, an educational necessity exists where the challenged practice serves the *legitimate* educational goals of the institution. *Wards Cove,* 490 U.S. at 659. In other words, the TEA must merely produce evidence that there is a manifest relationship between the TAAS test and a legitimate educational goal. *Teal,* 457 U.S. at 446. The Court finds that the TEA has met its burden.

The articulated goals of the implementation of the TAAS requirement are to hold schools, students, and teachers accountable for education and to ensure that all Texas students receive the same, adequate learning opportunities. These goals are certainly within the legitimate exercise of the State's power over public education. To determine whether the TAAS test bears a manifest relationship to these legitimate goals, the Court has considered carefully each of the test's alleged deficiencies—the overall effectiveness of the test, the cut score of the test, the use of the test as a requirement for graduation, the Plaintiffs' allegation that the test has resulted in inferior educational opportunities for minorities, and the alleged relationship between the test and student drop out scores.

A. Effectiveness

The Court finds that the TAAS test effectively measures students' mastery of the skills and knowledge the State of Texas has deemed graduating high school seniors must possess. The Plaintiffs provided evidence that, in many cases, success or failure in relevant subject-matter classes does not predict success or failure in that same area on the TAAS test. *See Supplemental Report of Dr. Walter Haney,* Plaintiffs expert, at 29-32. In other words, a student may perform reasonably well in a ninth-grade English class, for example, and still fail the English portion of the exit-level TAAS exam. The evidence suggests that the disparities are sharper for ethnic minorities. *Id.* at 33. However, the TEA has argued that a student's classroom grade cannot be equated to TAAS performance, as grades can measure a variety of factors, ranging from effort and improvement to objective mastery. The TAAS test is a solely objective measurement of mastery. The Court finds that, based on the evidence presented at trial, the test accomplishes what it sets out to accomplish, which is to provide an objective assessment of whether students have mastered a discrete set of skills and knowledge.

B. Cut Score

The Court has paid close attention to testimony in this case regarding the setting of the 70-percent passing standard for the TAAS test. In addition, the Court has carefully considered the scope of its own authority to address that issue. Ultimately, the Court concludes that the passing standard does bear a manifest relation to a legitimate goal.

Whether the use of a given cut score, or any cut score, is proper depends on whether the use of the score is justified. In *Cureton,* a case relied upon heavily by the Plaintiffs in this case, the court found that the use of an SAT cut score as a selection practice for the NCAA must be justified by some independent basis for choosing the cut score. *Cureton,* 37 F. Supp.2d at 708. In addition, the court noted that the NCAA had not validated the use of the SAT as a predictor for graduation rates. *Id.*

Here, the test use being challenged is the assessment of legislatively established minimum skills as a requisite for graduation. This is a conceptually different exercise from that of predicting graduation rates or success in employment or college. In addition, the Court finds that it is an exercise well within the State's power and authority. The State of Texas has determined that, to graduate, a senior must have mastered 70 percent of the tested minimal essentials.

In *Tyler v. Vickery*, 517 F.2d 1089 (5th Cir. 1975), the United States Court of Appeals for the Fifth Circuit noted two criteria for determining whether a standardized test is rationally supportable. *Tyler*, 517 F.2d at 1101. The relevant criterion here is whether the cut score is related to the quality the test purports to measure. *Id.* The court noted that a 70-percent cut score for bar passage "has no significance standing alone" but that it "represents the examiners' considered judgments as to minimal competence required to practice law." *Id.* The court finds that the 70-percent cut score for the TAAS test reflects similar judgments. *See Report of the State Board of Education Committee of the Whole, Work Session Minutes,* July 12, 1990. The Court does not mean to suggest that a state could arrive at any cut score without running afoul of the law. However, Texas relied on a field test data and input from educators to determine where to set its cut score. It set initial cut scores 10 percentage points lower, and phased in the 70-percent score. *See State Board of Education Minutes,* July 14, 1990. While field test results suggested that a large number of students would not pass at the 70-percent cut score, officials had reason to believe that those numbers were inflated. *See Work Session Minutes,* July 12, 1990. Officials contemplated the possible consequences and determined that the risk should be taken. The Court cannot say, based on the record, that the State's chosen cut score was arbitrary or unjustified. Moreover, the Court finds that the score bears a manifest relationship to the State's legitimate goals.

C. Use as a Graduation Requirement

The Court finds that the TEA has shown that the high-stakes use of the TAAS test as a graduation requirement guarantees that students will be motivated to learn the curriculum tested. While there was testimony that the test would be useful even if it were not offered as a requisite to graduation, the Court finds that there was no, or insufficient, evidence to refute the TEA's assertion that the use as a graduation requirement boosted student motivation and encouraged learning. In addition, the evidence was unrefuted that the State had an interest in setting standards as a basis for the awarding of diplomas. The use of a standardized test to determine whether those standards are met and as a basis for the awarding of a diploma has a manifest relationship to that goal.

D. Inferior Educational Opportunities

The Plaintiffs introduced evidence that, in attempting to ensure that minority students passed the TAAS test, the TEA was limiting their education to the barest elements. The Court finds that the question of whether the education of minority students is being limited by TAAS-directed instruction is not a proper subject for its review.[11] The State of Texas has determined that a set of knowledge and skills must be taught and learned in State schools. The State mandates no more than these "essential" items. Test-driven instruction undeniably helps to accomplish this goal. It is not within the Court's power to alter or broaden the curricular decisions made by the State.

E. Drop-out and Retention Rates

As discussed above, the Plaintiffs have presented credible evidence that the drop-out and retention rates among minority students in Texas give cause for concern. However, there is no credible evidence linking State drop-out and retention rates to the administration of the exit-level TAAS test. Expert Walter Haney's hypothesis that schools are retaining students in the ninth grade in order to inflate tenth-grade TAAS results was not supported with legally sufficient evidence demonstrating the link between retention and TAAS.

III. Equally Effective Alternatives

In considering whether the Plaintiffs have shown that there are equally effective alternatives to the current use of the TAAS test, the Court must begin with the State's articulated, legitimate goals in instituting the examination. Those goals are to hold students, teachers, and schools accountable for learning and for teaching, to ensure that all students have the opportunity to learn minimal skills and knowledge, and to make the Texas high school diploma uniformly meaningful. Further, as discussed more fully above, the State has set a standard for mastery of 70 percent of the items tested, and the Court has held that this standard is legitimate.

Plaintiffs did offer evidence that different approaches would aid the State in measuring the acquisition of essential skills. Among these approaches was a sliding-scale system that would allow educators to compensate a student's low test performance with high academic grades or to compensate lower grades with outstanding test scores. However, Plaintiffs failed to present evidence that this, or other, alternatives could sufficiently motivate students to perform to their highest ability. In addition, and perhaps more importantly, the present use of the TAAS test motivates schools and teachers to provide an adequate and fair education, at least of the minimum skills required by the State to all students. *See Debra P. II,* 730 F.2d at 1416. The Plaintiffs produced no alternative that adequately addressed the goal of systemic accountability.

DUE PROCESS

In order for a court to find a due process violation, it must first find that a plaintiff has a protected interest—either property or liberty—in what the State seeks to limit or deny. *See Michael H. v. Gerald D.*, 491 U.S. 110, 121 (1989) (substantive due process, liberty interest); *Ewing*, 474 U.S. at 222 (substantive due process, property interest); *Ewing*, 474 U.S. at 229 (procedural due process, property interest). The Court has previously found, and reiterates here, that the State of Texas has created a protected interest in the receipt of a high school diploma. *See* TEX. EDUC. CODE § 25.085(b); *id.* at § 4.002; *id.* at § 28.025(a)(1); *Debra P.*, 644 F.2d at 403-404.

The Due Process Clause has two aspects—procedural and substantive. *Ewing*, 474 U.S. at 229. On the procedural side, the law demands that a state provide, at a minimum, notice and an opportunity to be heard before it deprives citizens of certain state-created protected interests. *Frazier v. Garrison I.S.D.*, 980 F.2d 1514, 1529 (5th Cir. 1993). On the substantive side, the law holds that some rights are so profoundly inherent in the American system of justice that they cannot be limited or deprived arbitrarily, even if the procedures afforded an individual are fair. *Ewing*, 474 U.S. at 229, *Robertson v. Plano City*, 70 F.3d 21, 24 (5th Cir. 1995). The use of a standardized test as a graduation requirement can implicate both procedural due process concerns and substantive due process concerns. *Debra P.*, 644 F.2d at 404.

The United States Court of Appeals for the Fifth Circuit has held that a state cannot impose a standardized test as a graduation requirement without giving its students the procedural protection of adequate notice that such will be the use of the test. *Id.* at 404. In addition, the Fifth Circuit has suggested a *substantive* component to a student's rights where a state attempts to condition a diploma on standardized test scores: a state may not impose an examination where such imposition is arbitrary and capricious or frustrates a legitimate state interest or is fundamentally unfair, in that it encroaches upon concepts of justice lying at the basis of our civil and political institutions. *Id.* The United States Supreme Court has suggested that a state's educational determinations may be invalid under a substantive due process analysis where they reflect a "substantial departure from accepted academic norms as to demonstrate that the person or committee responsible did not actually exercise professional judgment." *Ewing*, 474 U.S. at 225. The Court has evaluated the use of the TAAS examination under each of these formulations and finds that it does not violate the due process rights of Texas students, minority or majority.

A test that covers matters not taught in the schools is fundamentally unfair. *Debra P.*, 644 F.2d at 404. The Court finds, however, that the TAAS exit-level test meets currently accepted standards for curricular validity. In other words, the test measures what it purports to measure, and it does so with a sufficient degree of reliability. In addition, all students in Texas have had a reasonable opportunity to learn the subject matters covered by the exam. The State's efforts at remediation and the fact that students are given eight opportunities to pass the examination before leaving school support this conclusion. *Debra P. II*, 730 F.2d. at 1411.

The Court also finds that the Plaintiffs have not demonstrated that the TAAS test is a substantial departure from accepted academic norms or is based on a failure to exercise professional judgment. Certainly, there was conflicting evidence at trial regarding whether the test, as used, is appropriate. However, there was no testimony demonstrating that Texas has rejected current academic standards in designing its educational system. Educators and test-designers testified that the design and the use of the test was within accepted norms.

The Court, in reaching this conclusion, has considered carefully the testimony of Plaintiffs' expert, Dr. Martin Shapiro, demonstrating that the item-selection system chosen by TEA often results in the favoring of items on which minorities will perform poorly, while disfavoring items where discrepancies are less wide. The Court cannot quarrel with this evidence. However, the Court finds that the Plaintiffs have not been able to demonstrate that the test, as validated and equated, does not best serve the State's goals of identifying and remediating educational problems. Because one of the goals of the TAAS test is to identify and remedy problems in the State's educational system, no matter their source, then it would be reasonable for the State to validate and equate test items on some basis other than their disparate impact on certain groups. In addition, the State need not equate its test on the basis of standards it rejects, such as subjective teacher evaluations.

In short, the Court finds, on the basis of the evidence presented at trial, that the disparities in test scores do not result from flaws in the test or in the way it is administered. Instead, as the Plaintiffs themselves have argued, some minority students have, for a myriad of reasons, failed to keep up (or catch up) with their majority counterparts. It may be, as the TEA argues, that the TAAS test is one weapon in the fight to remedy this problem. At any rate, the State is within its power to choose this remedy.

As the court has stated in prior orders, it would be fundamentally unfair to punish minority students for receiving an unequal, state-funded education.[12] In other words, it would violate due process if the TAAS test were used as a vehicle for holding students accountable for an educational system that failed them. The Court concludes, however, that the TAAS test is not used in such a manner.

The Court has considered this question carefully. Texas's difficulties in providing an equal education to all its students are well-documented. It is only in the recent past that efforts have been made to provide equal funding to Texas public schools. Several schools in the state remain under desegregation orders. These facts cannot be ignored.

The Court finds, however, after listening to the evidence at trial, that the TEA would agree with the proposition that unequal education is a matter of great concern and must be eradicated. The Court has determined that the use and implementation of the TAAS test does identify educational inequalities and attempts to address them. *See Debra P. II,* 730 F.2d at 1415 (remedial efforts help dispel link between past discrimination and poor performance on standardized test). While lack of effort and creativity at the local level sometimes frustrate those attempts, local policy is not an issue before the Court. The results of the TAAS test are used, in many cases quite effectively, to motivate not only students but schools and teachers to raise and meet educational standards.

CONCLUSION

ACCORDINGLY, the Court finds that the TAAS exit-level examination does not violate regulations enacted pursuant to Title VI of the Civil Rights Act of 1964. While the TAAS test does adversely affect minority students in significant numbers, the TEA has demonstrated an educational necessity for the test, and the Plaintiffs have failed to identify equally effective alternatives. In addition, the Court concludes that the TAAS test violates neither the procedural nor the substantive due process rights of the Plaintiffs. The TEA has provided adequate notice of the consequences of the exam and has ensured that the exam is strongly correlated to material actually taught in the classroom. In addition, the test is valid and in keeping with current educational norms. Finally, the test does not perpetuate prior educational discrimination or unfairly hold Texas minority students accountable for the failures of the State's educational system. Instead, the test seeks to identify inequities and to address them. It is not for this Court to determine whether Texas has chosen the best of all possible means for achieving these goals. The system is not perfect, but the Court cannot say that it is unconstitutional. Judgment is GRANTED in favor of the Defendants, and this case is DISMISSED.

SIGNED AND ENTERED this 7th day of January 2000.

[signature]
EDWARD C. PRADO
UNITED STATES DISTRICT JUDGE

Notes

[1] This suit is also brought individually by nine Texas students who did not pass the TAAS exit-level examination prior to their scheduled graduation dates. Those students who actually testified request that their respective school districts issue their diplomas. Consistent with this Order, that request is denied. Those students who did not appear to testify—Melissa Marie Cruz, Michelle Marie Cruz, and Jocqulyn Russell—are dismissed from the case for failure to prosecute.

[2] The Court read and heard with interest the conclusions of Plaintiff's expert Amilcar Shabazz

on this subject. *See Report of Dr. Amilcar Shabazz,* Plaintiff's expert, at 11-12. Shabazz rejects the argument that offering focused remedial efforts to students who do not pass the TAAS helps eradicate the effects of past discrimination. A student who fails the test does not graduate. A student who has been remediated and finally passes the test has only passed a test, not necessarily received an adequate education. The Court notes in response that its authority to determine what constitutes an "adequate" education is extremely limited.

[3] Of course, there are generalizations. The Court recognizes that students in districts with relatively greater resources have failed the TAAS examination.

[4] The Court does not suggest that the psychometricians who testified on behalf of the TEA reject the notion that a test's effects should be fair. Rather, they view the system in place, which provides wholly objective assessment, as the best way to ensure fairness. In addition, Defendant's expert Dr. Susan Phillips noted that careful scrutiny is given to test items that are identified as having large differences between the performances of minority and majority students. *See Report of Dr. Susan Phillips,* Defendants' expert, at 3.

[5] Any finding of fact more appropriately characterized as a conclusion of law may be considered as such.

[6] In 1998-1999, the Texas Essential Knowledge and Skills (TEKS) replaced the Essential Elements.

[7] The Four-Fifths Rule finds an adverse impact where the passing rate for the minority group is less than 80 percent of the passing rate for the majority group. 29 C.F.R. ' 1607.

[8] The *Shoben* formula seeks to assess the statistical significance of observed numerical disparities by determining differences between independent proportions. *See Frazier v. Consolidated Rail Corp.,* 851 F.2d 1447, 1450 n.5 (D.C. Cir. 1988).

[9] Any conclusion of law more appropriately characterized as a finding of fact may be considered as such.

[10] As noted elsewhere, the TEA has suggested that this regulation has been limited to its constitutional dimensions (i.e., to a requirement that a plaintiff show discriminatory intent) by the United States Supreme Court, in *United States v. Fordice,* 505 U.S. 717 (1992). The Court acknowledges the dicta to which the TEA refers. *See Fordice,* 505 U.S. at 732. However, the Courts notes that other courts have not held that the disparate impact analysis under 34 C.F.R. § 100.3 has been abrogated. *See Cureton,* 37 F. Supp.2d at 697 (collecting cases); *Graham v. Tennessee Secondary Sch. Athletic Assoc.,* No. 1:95-cv-044, 1995 WL 115890, at * 12 (E.D. Tenn. Feb. 20, 1995) (joining other courts in maintaining disparate impact claim after *Fordice*). It is this Court's duty to *apply* the law, as near as it is able, and only to *predict* what the law will be when absolutely necessary. *See Charles J. Cooper, Stare Decisis: Precedent & Principal in Constitutional Adjudication,* 73 CORNELL L. REV. 401 at n.6 (1988).

[11] Of course, upon a showing of intentional discrimination, such a claim would implicate the Equal Protection Clause of the Fourteenth Amendment. However, the Court has already held that Plaintiffs have offered no proof of intent in this case.

[12] In *Debra P. II,* the United States Court of Appeals for the Fifth Circuit articulated this concern in equal protection terms, reiterating the proposition that an educational system still suffering from the effects of prior discrimination cannot classify students based on race unless that classification can be shown either not to be a result of prior discrimination or that it will remedy such discrimination. *See Debra P. II,* 730 F.2d at 1411. This Court has dismissed the Plaintiffs equal protection claim. Nonetheless, the Court has stated, and emphasizes again here, that it would be a due process violation to impose standards on minority students whose failure to meet those standards is directly attributable to state action.

Accountability Is Overdue

Testing the Academic Achievement of Limited-English Proficient (LEP) Students

Rosalie Pedalino Porter, Ed.D.

Since the 1960s, the United States has received the highest number of new arrivals in the nation's history—legal and illegal immigrants, migrants, and refugees. Consequently, U. S. public schools have seen a rapidly increasing enrollment of immigrant children, and of native-born children of immigrant parents, who have little or no fluency or literacy in English. Providing these 3.5 million children with an educational opportunity equal to that of English speakers is the challenge, and legislation, court decisions, and education policies have been attempting to meet this challenge for the past 30 years.

It was a Texas senator, Ralph Yarborough, who filed the first federal legislation to address the problem: the Bilingual Education Act of 1968, Title VII of the Elementary and Secondary Education Act. The goal at the beginning was to help poor Mexican-American children learn English, although this was later expanded to include non-English speaking children of any language background. Yarborough said at the time, "It is not the purpose of the bill to create pockets of different languages through the country...but just to try to make those children fully literate in English" (Chavez, 11-12). Starting with Massachusetts in 1971, state laws were enacted that required bilingual schooling for a few years to help children overcome the language barrier to an equal education. The U.S. Supreme Court in its *Lau v. Nichols* decision in 1974 (*Lau*) declared that non-English speaking children have a right to special help.

> There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand English are effectively foreclosed from any meaningful education... Teaching English to the students of Chinese ancestry who do not speak the language is one choice. Giving instruction to the group in Chinese is another. There may be others.

The decision in *Castaneda v. Pickard,* (648 F. 2nd 989, Fifth Circuit, 1981) established a three-pronged test for determining whether a school district is taking appropriate action to overcome language barriers, as follows:

1. The school district is pursuing a program informed by an educational theory recognized as sound by some experts in the field.

2. The programs and practices actually used by a school system are reasonably expected to implement the educational theory adopted by the school, that sufficient resources are provided (i.e., trained teachers, textbooks).
3. After a sufficient length of time, proper evaluation of the special program shows results indicating that language barriers are actually being overcome (Rebell and Murdaugh, 1992, p. 365).

It is the third *Castaneda* standard that brings accountability into the entire national effort to help limited-English students. It requires that at some point, in a few years at most, there must be clear evidence that students have benefited from this special help, that in fact they have progressed academically both in learning the English language and in their ability to learn school subjects taught in English.

Texas is perhaps the best example of what can be accomplished in a relatively short period of time in improving student performance on objective measures of curriculum and skills taught in all schools. Not only has performance improved across the board for all students since the statewide testing program began in 1985, but minority students—African American and Hispanic students—have achieved the highest rates of improvement and are gradually closing the performance gap with their white classmates. In the most recent 10th-grade test, spring 1999, 95 percent of white students passed the test compared with 84 percent of Hispanic and African American students—a commendable result compared to minority student achievement in other states, such as Massachusetts and New York, for example.

Suffice it to say that the amount of human capital invested—in developing curriculum standards, training teachers, developing and annually reviewing and modifying tests, and in collecting and reporting student performance data—is remarkable and presents a useful model for the rest of the country. Although the Texas Assessment of Academic Skills (TAAS) is administered in grades 2-8 and in grade 10, I am restricting my discussion to the 10th-grade only, as it is the "high stakes" test that is challenged in the *G. I. Forum v. Texas Education Agency* lawsuit.

I am confining my remarks further to the sub-group of Hispanic students that is defined as LEP. It is important to understand the distinction. The majority of Texas school children of Spanish-speaking families are native-born, English-language speakers when they enter the schools. Those labeled "LEP" are children of immigrant or migrant families more recently arrived in Texas. For this particular group of children, there are many considerations that affect their rate of English language learning and academic progress as it is reflected in their test scores: age at arrival in the U. S., previous level and quality of schooling in their land of origin, educational level of the parents, economic status, whether the family moves often (especially common for migrant worker families), type of special program in which children are enrolled (Spanish bilingual instruction, English as a Second Language, or no special program).

It matters greatly, for instance, if an LEP child started school in Texas in kindergarten and with some knowledge of English and with 11 years of schooling before taking the 10th-grade exams, or if the student arrived in Texas at the eighth- or ninth-grade level with few years of schooling in his or her native land and no fluency in English at all. However, this kind of data does not appear on the report summarizing test scores. Performance is reported in groups by ethnic category and, for language minority children, under the further headings Migrant, Limited-English Proficient, Bilingual Program Participant, and ESL (English as a Second Language) Program Participant.

By charting the progress of LEP students since the 10th-grade test has been required for high school graduation, it is useful to compare the percent who met the minimum expectations on all tests (reading, mathematics and writing) in 1994 and 1999, as illustrated in Table 1 on page 109. In 1994 a total of 187,618 students were tested at that grade level of whom 52 percent (including LEPs but not students in special education) met the minimum expectations on all tests taken. In 1999, 213,959 took the 10th-grade tests and 78 percent were successful in all tests taken. Clearly, more students are participating in the assessments and more are at least meeting minimum expectations for high school graduation. The record for LEP students as a separate group is not as inspiring, but there is steady improvement documented.

The number of limited-English students participating in the 10th-grade test has increased from 19,167 to 23,120, and the percentage of students passing all three parts of the test has more than doubled in this five-year period. What is not reported is how many of the students in the three categories who did not score at the minimum expectation level took advantage of the remedial classes offered and of the multiple opportunities to retake the test. Also, the reason for separately listing the three categories is not clear and needs fuller explanation. All the students in these three categories are limited-English to some degree. Some are participating in bilingual classes, some in ESL classes.

In the states with large enrollments of LEP students, evaluation of LEP student achievement has been very little attended to in the past 30 years. Two representative examples, California and Massachusetts, serve to illustrate this lack of accountability. California enrolls 43 percent of all LEP students in the country, 1.4 million children who start school without the ability to do regular classroom work in English. *Meeting the Challenge of Language Diversity,* (Berman et al., 1992), is the first statewide report on the outcomes of bilingual education programs, and it reveals a serious lack of consistent student testing or data collection by the California State Department of Education. Conclusion 6 of the report asserts: "California public schools do not have valid and ongoing assessments of the performance for students with limited proficiency in English. Therefore, the state and the public cannot hold schools accountable for LEP students achieving high levels of performance" (Rossier, 1995, p. 46). It is reasonable to question this stunning admission by asking, if the schools are not accountable for student learning, then who is?

In 1998, California instituted the Standardized Testing and Reporting (STAR) program that requires all students to participate at every grade level from second to 11th grade, including LEP students. For those LEP students who have been in California schools fewer than 12 months, a comparable test may be taken in the native language, if available. At this writing, standardized tests are available only in Spanish. Finally, it is now possible to identify the students, schools and districts that need improvement at particular grade levels and in certain subject areas, so that appropriate additional resources can be provided for those needs. After two test administrations, California reports improved performance for limited-English students at every grade level although the average performance is disappointingly low. For example, the reading scores for LEP second-graders across the state rose from the 19th to the 23rd percentile, and all students tested at that grade level increased scores from the 39th to the 43rd percentile (Porter, 1999).

Massachusetts, the first state to enact legislation on bilingual schooling in 1971, had equally shirked its legal responsibility to document the progress of LEP students until very recently. Not one recognized research study evaluating bilingual programs has been published in this state. *Striving for Success,* a statewide survey published in 1994 reported,

> The Commission found that adequate and reliable data has never been collected that would indicate whether or not bilingual programs offer language minority pupils a superior educational option. This report strongly endorses the 1993 Education Reform Act's emphasis on accountability of educational outcomes for all pupils, including the development of appropriate assessments of pupils in bilingual programs and the collection of data specific to bilingual pupils (Massachusetts Bilingual Education Commission Report, p. 2).

Massachusetts is now one of 26 states that not only mandate annual testing of students but also require a passing grade on the 10th-grade assessment for high school graduation. Passing the 10th-grade test will be essential for all students in Massachusetts, starting in 2003, 10 years after the Education Reform Act began financing the development of curricular frameworks in all subjects and related tests to evaluate student learning.

The legislature has allocated generous new education funding every year, especially to urban districts with high enrollments of minority students from low-income families. The Massachusetts Comprehensive Assessment System (MCAS) is administered to fourth-, eighth-, and 10th-graders. After only two test administrations, early results show these highlights:

- Test participation is high with 96 percent of all students being tested, including students with disabilities and limited-English students.
- The tests on which the highest percentage of students performed at the two top levels, *Advancea* and *Proficient:*
    - Grade 4 Science and Technology - 56 percent
    - Grade 4 Mathematics - 36 percent
    - Grade 8 English Language Arts - 56 percent
    - Grade 10 English Language Arts - 34 percent
- The tests on which the highest percentage of students performed at the *Failing* level:
    - Grade 8 History and Social Science - 49 percent
    - Grade 8 Mathematics - 40 percent
    - Grade 8 Science & Technology - 45 percent
    - Grade 10 Mathematics - 53 percent
- Especially disappointing are results on the 10th-grade tests for students classified as LEP, although these students are not required to take the MCAS tests in English until they have been in U. S. schools three years or longer. Percent of LEP students scoring at the *Failing* level in English Language Arts 66 percent; in Mathematics 92 percent; and in Science and Technology 80 percent.
- In both 1998 and 1999 students at grade 4 had the highest average scaled scores overall and the lowest percentage of students at the *Failing* level (Massachusetts Comprehensive Assessment System, pp. 3-4).

Although these results indicate substantial room for improvement, they are by no means unusual. When statewide assessments of academic performance are first employed, the results may be less satisfactory than expected. New York state, for example, is at an early stage of measuring student achievement with new, more rigorous, tests. New York state reported more than half of fourth graders failed the new English test and 33 percent were below standard in mathematics. At the eighth grade level, 52 percent were below standard in reading and 62 percent in mathematics (Hartocollis, 1999, pp. 1, 14).

One of the major reasons for the low percentage of Hispanic high school graduates, both in Texas and across the country, is the high dropout rate for this population. In spite of special programs for Hispanic students, the dropout rate has not appreciably improved nationally over the past 25 years. According to a recent report, nationally the Hispanic dropout rate has remained between 30 percent and 35 percent during this period, two and a half times the rate for African Americans and three and a half times the rate for white non-Hispanics. (Hispanic Dropout Project, 1999, p. 5) This dropout disproportion is part of the problem in Texas as well. The Texas Education Agency claims that 2.3 percent of the state's Hispanic students drop out of school each year between grades 7 and 12, compared to a .9 percent rate for white students. Consequently, although Hispanics make up 37 percent of the state's students, they only account for 29 percent of its high school graduates (Kronholz, 1999, p. 20).

On the central question of this lawsuit—whether high school students should be expected to demonstrate competency in reading, writing and mathematics on an objective measure such as the 10th-grade TAAS test in order to obtain a high school diploma—I am firmly convinced of the position of the Texas Education Agency that this testing program is urgently needed. In my professional opinion, it is sound educational policy to require one objective, uniform measure of student achievement as a prerequisite for high school graduation, an assessment closely based on the material taught in the schools. To suggest that students should graduate without demonstrating minimal knowledge and skills on a uniform measure is not acceptable for the current requirements of the technological/information age job market or for pursuing higher education. Delia Pompa, (as cited in Porter, 1994) director of the Office of Bilingual Education and Minority Languages Affairs in the U.S. Department of Education, commented pointedly on the need for LEP students to be held to reasonable learning standards and assessments: "I'm not sure it's O.K. for our kids to dance out something where other kids have to write on a subject to show mastery" (p. 44).

Exempting whole groups of students from statewide assessments on the expectation that they will not perform adequately is unfair to the students who are excluded as well as to their classmates. It has been my experience as a teacher and as a program administrator that the majority of English language learners want to be included in the same educational and testing programs as native English speakers and that they feel demeaned when they are left out. A policy of separating language minority students, many of whom are native born, from the rest of the student population when the TAAS is administered is more likely to stigmatize and negatively impact the self-esteem of these students than is their inclusion in the tests.

In the case of minority students and especially LEP students, the TAAS program reported the urgent need for extraordinary efforts to be directed to these populations. Texas has well documented the educational improvements implemented and the steady growth in successful performance on state tests. A past history of discrimination against Mexican-American and

African-American children is not justification for holding these students to lower standards. Dr. Jose Cardenas, a witness for the plaintiffs in the Texas case, has stated, nevertheless, that Texas has done much to eliminate discriminatory practices in the education of minority students in the past two decades. Maintaining rigorous standards and high expectations for minority students requires that periodic assessments of each student's progress be conducted and reported. The useful data collected annually not only play a part in improving teaching and learning but are used to modify the TAAS program itself.

In my 25 years of work in the bilingual education field, one of the major themes stressed continually to teachers and administrators is the importance of communicating to our students that we have high expectations for their ability to meet the same standards as other students. We expect them to reach high levels of achievement with our help. Discontinuing the process of accountability for Limited-English Proficient students in Texas would be a disservice to a group of students whose academic progress has not been monitored heretofore in a consistent, longitudinal manner. As an expert witness in this case on behalf of the Texas Education Agency, I applaud Judge Edward C. Prado's ruling on January 7, 2000, that the TAAS "is not perfect, but the Court cannot say that it is unconstitutional." He recognizes that the test "does not perpetuate prior educational discrimination.... Instead, the test seeks to identify inequities and to address them." On February 8, 2000, the Mexican American Legal Defense and Education Fund (MALDEF) announced that it will not be appealing the ruling of Judge Prado (MALDEF announces, p. 9).

This is the crux of the matter: without a statewide, annual, consistent, universally applied program of assessment, the next logical step of improving student achievement cannot be accurately addressed. Had Judge Prado ruled otherwise, it would have set an unfortunate precedent for other states with large numbers of LEP students where accountability is still in the early stages.

Certainly there are many forms of assessment that are valuable, including portfolios, classroom work, and teacher evaluations. However, these evaluations are not consistent from school to school or district to district. At some point, and the 10th-grade tests of basic skills is, in my opinion, the time for this assessment, students must be able to demonstrate on a universally applied measure that they can read, write, and do mathematics at least at a minimal level if their high school diploma is to have any validity.

[This article was first published in Applied Measurement in Education, *vol. 13, no. 4, October 2000, and appears here by permission of the author and Lawrence Erlbaum, Publishers.]*

Table 1.
LEP Students Meeting Minimum Expectations on All Tests

	1994	1999		
	# Tested	% Passing All Tests	# Tested	% Passing All Tests
LEP	11,127	14%	12,903	31%
Bilingual Participants	95	18%	50	35%
ESL Participants	7,945	9%	10,167	27%

(Chart compiled by the author from data reported by Texas Education Agency, December 30, 1999)

References

Berman, P., Chambers, J., Gandara, P., McLaughlin, B., Minicucci, C., Nelson, D., Olsen, L. and Parrish, T. (1992). *Meeting the challenge of language diversity: An evaluation of programs for pupils with limited proficiency in English.* Berkeley, Calif.: BW Associates.

Castaneda v. Pickard, 648 F. 2nd 989 (5th Circuit, 1981).

Chavez, L. (1991). *Out of the barrio: Towards a new politics of Hispanic assimilation.* New York: Basic Books.

G. I. Forum et al. v. Texas Education Agency et al. (1997). Civil Action No. SA-97-CA-1278-EP, U. S. District Court, Western District of Texas, San Antonio Division.

Hartocollis, A. Eighth graders fail two New York tests. *New York Times,* Nov. 6, 1999, pp. A1, A14.

Hispanic Dropout Project. (February 1998). *No more excuses: The final report of the Hispanic dropout project.* Washington, D.C.: U. S. Department of Education.

Kronholz, J. Texas student exams face political and judicial tests. *Wall Street Journal,* Dec. 23, 1999, p. 20.

Lau v. Nichols, 414 U.S. 563 (1974).

MALDEF announces that it will not be appealing Judge Prado's ruling. (Feb. 8, 2000). *The Washington Post* (from news service), p. 9.

Massachusetts Bilingual Education Commission. (1994). *Striving for success: The education of bilingual pupils.* Boston: Executive Office of Education.

Massachusetts Comprehensive Assessment System: Report of 1999 state results. Malden, Mass.: Massachusetts Department of Education.

Porter, R. P. (May 18, 1994). Goals 2000 and the bilingual student. *Education Week,* 13 (34), 44.

_________. (1999, December/2000, January). The benefits of English immersion. *Educational Leadership,* 57 (4), 52-56.

Rebell, M. A. and Murdaugh, A. W. (1992) National values and community values: Equal opportunity for limited-English proficient students. *Journal of Law & Education,* 21(3), pp. 335-380.

Rossier, R. E. (1995). A critique of California's evaluation of programs for students of limited-English proficiency. *READ Perspectives,* 2 (1), pp. 27-54.

Recognizing Successful Schools for High-Achieving, Low-Income Students

The "No Excuses" Campaign

Robert E. Rossier, Ph.D.

> Nationwide, 58 percent of low-income fourth-graders in the United States cannot read. Sixty-seven percent of low-income inner-city eighth-graders cannot meet basic math standards for their grade level. Inner-city blacks and Latinos have suffered the worst because of this failure to teach basic skills. This national tragedy does not have to be. The seven Salvatori winners show that all children can excel academically regardless of race, income level, or family background. All seven of their schools score at or above the 65th percentile on nationally norm-referenced exams even though 75 percent or more of their students qualify for the free or reduced-price lunch."
>
> From the Introduction (p. 2)
> *No Excuses*

The family of the late Henry Salvatori emigrated to the United States from Italy when he was a young child. Growing up, he took advantage of the opportunities he found here by securing an education and launching a career in petroleum geology. He eventually founded what was to become one of the leading petroleum-exploration companies in the world.

Because of his successful experience, Salvatori decided that, through philanthropy, he would help open the door to opportunity for others. One of his initiatives has been the "No Excuses Campaign," a national effort directed by the Heritage Foundation in Washington, D.C., to assist the public schools in bettering academic achievement for all children, whatever their race, ethnicity, or socioeconomic level. The unifying theme of this campaign is: "There is no excuse for the academic failure of most public schools serving poor children" (*No Excuses*, p. ii).

The No Excuses Campaign focuses on school principals, those individuals who direct our schools and thus determine, in large part, a school's relative success. The 1999 Salvatori Prize for American Citizenship was awarded to seven school principals who have demonstrated that "our nation's poorest schools can become centers of academic excellence." Geographically, the schools of the seven honorees are diverse, representing great cities across the continent:, Chicago, Detroit, Houston, three New York City schools, and Inglewood, Calif., a suburb of Los Angeles. Irrespective of their locations, the students in each of these schools have shown superior academic achievement, scoring well above the national average on nationally norm-referenced tests even though three-quarters or more of the students come from poverty-level homes (p. 2).

Common Elements of Prize-Winning Schools

What magic do these principals have that enables them to transform their schools into vibrant centers of learning? According to the Heritage Foundation, directors of the No Excuses Campaign, there are seven common elements that must be present in high-performing, high-poverty schools:

1. Principals must be free to make decisions critical to the efficient operation of their schools and instructional programs—effective principals decide how to spend their money, whom to hire, and what to teach.
2. Principals should use measurable goals to establish a culture of achievement—once a principal sets a clear vision for the school, every teacher has to be held personally responsible for enforcing it.
3. Master teachers bring out the best in a faculty, and effective principals are discriminating in recruiting the very best teachers they can find and in designing their curriculum around the strengths and expertise of their staff.
4. Rigorous and regular testing leads to continuous student achievement—regular tests at all levels and in all areas ensure that teaching and learning of the prescribed curricula are taking place in every classroom.
5. Achievement in the school is the key to positive discipline—self-control, self-reliance, and self-esteem anchored in achievement are the means to success.

6. Principals must work actively with parents to make the home a center of learning—effective principals establish contracts with parents to support their children's efforts to learn.
7. The effort involved in achievement creates ability. Time on task is the key to progress in school. Effective principals demand hard work of their students and provide for extended days, after-school programs, summer programs—none wastes time.

Seven Honorees

The seven winners of the Salvatori Prize are listed here with a brief description of their schools, of the level of student achievement in reading and math, and of the particular standardized tests used by the schools.

Irwin Kurz
P.S. 161–The Crown School, Brooklyn, N.Y.
1,342 Students; 98 percent low income
1998 Average Test Scores, Grades 3-8:
National Percentile in Reading: 71
National Percentile in Math: 78
Grades K-8
California Test of Basic Skills (CTBS) and
California Achievement Test-5 (CAT-5)

Gregory Hodge
Frederick Douglass Academy, New York City
1,030 Students; 81 percent low income
1998 Average Test Scores, Grades 7-8:
National Percentile in Reading 73
National Percentile in Math 81
Grades 7-12
CTBS and CAT-5

Michael Feinberg
KIPP Academy, Houston, Texas
270 Students; 95 percent low income
1998 Average Test Scores, Grades 5-9:
National Percentile in Reading 61
National Percentile in Math 81
Stanford-9 Achievement Test

David Levin
KIPP Academy, Bronx, N.Y.
223 Students; 95 percent low income

1998 Average Test Scores, Grades 5-8:
National Percentile in Reading 69
National Percentile in Math 81
CTBS and CAT-5

Nancy Ichinaga
Bennett-Kew Elementary School, Inglewood, Calif.
836 Students; 78 percent low income
1998 Average Test Scores, Grades 2-5:
National Percentile in Reading 58
National Percentile in Math 67
Grades K-5
Stanford-9 Achievement Test

Helen DeBerry
Earhart Elementary, Chicago
265 Students; 82 percent low income
1998 Average Test Scores, Grades 1-6:
National Percentile in Reading 70
National Percentile in Math 80
Grades PK-6
Iowa Test of Basic Skills

Ernestine Sanders
Cornerstone Schools Association, Detroit
625 Students; 75 percent low income
1998 Average Test Scores, Grades 1-8
National Percentile in Reading 65
National Percentile in Math 51
Grades PK-8
Stanford-9 Achievement Test

A Special Case

Of the seven honorees, Nancy Ichinaga of the Bennett-Kew Elementary School in California, is singled out for a more detailed account because her school has a high enrollment of English language learners (formerly referred to as Limited-English Proficient [LEP] students) that is of special interest to readers of *READ Perspectives*. Ichinaga is also recognized for her courage and tenacity. On two separate occasions she led a coalition of parents and teachers that successfully fought the California educational establishment over the question of instructional approaches to be used in her school.

Ichinaga is unwavering in her conviction that the primary mission of the school is to help children to become literate and that this task should begin in kindergarten. She and her staff are firmly committed to a reading program that has "a systematic decoding component" (p. 23). It was her dedication to a phonics approach that brought about a confrontation in 1986 with the California State Curriculum Commission. The Commission backed a "whole language approach" for reading instruction in the state's schools and did not allow any deviation from that methodology. For that reason, the Commission decided to withhold state funds that Bennett-Kew needed to buy textbooks for its phonics reading program. Declaring war, Ichinaga and her students' parents mounted a massive letter-writing campaign that forced the Commission to back down and allow the school's phonics texts to be placed on the state's list of approved books.

It was inevitable that this feisty principal would clash with state authorities on another question of great importance to California's system of public education. Bilingual education was introduced into the state's schools in 1976 as a proposed solution to the problems occasioned by the influx into public school classrooms of hundreds of thousands of children with a limited knowledge of the English language.

By the early 1990s, 50 percent of Bennett-Kew's students were Hispanic with 30 percent of the entire school enrollment consisting of children who are English language learners. Despite the presence in the school of this large group of students who spoke little or no English, Ichinaga established a special English-based instruction program for these students that closely approximated the curriculum for the school as a whole. With certain modifications, Bennett-Kew's program adheres to Ichinaga's fundamental beliefs about how these students will learn English: Only English is used for instruction, English language learners are not segregated but rather are integrated with native English speakers.

Every student in the school, whether English speaking or not, is promoted to the next grade depending on his or her ability to meet standards of achievement that are clearly defined for each grade level. The promotion policy applies to all grades but especially to the kindergarten level. Ichinaga believes that the key to a successful reading program is to begin the drive for English reading mastery in kindergarten.

While the Bennett-Kew program for limited-English students follows the regular school curriculum in large part, there are several differences: All English language learners receive lessons in English as a Second Language (ESL) for 30 minutes each day, and Spanish-speaking instructional aides are in the classroom in the lower grades.

In a statement describing the organization and history of the Bennett-Kew program for English language learners, Ichinaga spells out the philosophy upon which the program is based:

> We have found that one of the most effective ways of teaching our Hispanic children to become English speakers is to immerse them in English from their first day in kindergarten. We have also found that in order to teach them successfully we need to use what is known as a 'total physical response' approach to language acquisition. Kindergarten children, regardless of their language backgrounds, learn through music and movement, rhythm and rhyme, finger plays, interacting with teacher-read books with lots of pictures....Through daily phoneme awareness activities and systematic phonic instruction, our children learn to read simple words in kindergarten. (Ichinaga and Schieldge, p. 2).

A *Los Angeles Times* article in 1992 spotlighting the success of Bennett-Kew and another Inglewood school, Kelso Elementary, triggered a reaction from the state's bilingual education bureaucracy. (Fuetsch) First, the two schools were accused of violating the civil rights of their Hispanic children and then, several months later, they were visited by a state compliance team. The team promptly charged the schools with not complying with the mandated state bilingual program and threatened the Inglewood School District with the withholding of $7 million in federal funds. The district was given a year to reach compliance with state rules. This drastic punishment was to be inflicted despite the schools' record of meeting the fundamental objectives of the state program: having academic achievement levels equal to or better than the state average for all students; and, having a redesignation rate from "Limited" to "Fluent English Proficient" status superior to that of the rest of the state's schools that used the same exit criteria. (Under California state guidelines, English language learners have been labeled "Limited-English Proficient" or LEP. Once these students have mastered the speaking, reading, and writing of the English language, they are then characterized as "Fluent English Proficient" or FEP, and "redesignated" or exited from their special program and assigned to regular mainstream classroom instruction in English.)

Once again, Ichinaga decided to fight. Her school staff asked each parent of a limited-English student to sign a request that their child be taught in English, not Spanish. The parent's preference, in writing, for an English instruction program did not carry much weight with the team, and the district remained out of compliance for a time. A year after the parent requests were turned over to the California Department of Education, a member of the state compliance team questioned the authenticity of the parent requests. Even after the team member was allowed to interview the parents, she remained skeptical of the sincerity of their preference for English language instruction for their children.

Finally, in 1996, the Department of Education granted the Inglewood schools a waiver from the obligation to teach subject matter in Spanish. It was clear that the excellent performance of Bennet-Kew and other schools in Inglewood with equally high test scores and redesignation rates had backed the Department into a corner. In the year that the waiver was granted, the statewide redesignation rate for English language learners was 6 percent, the same as it had been for several years, while the Bennett-Kew rate was 40 percent.

The struggle by the Inglewood schools for local program choice in educating language minority children finally ended in June 1998 with the passage of Proposition 227, the initiative that mandates English language instruction for all students in California's public schools. In the first year since Proposition 227 has been implemented, there is clear evidence that many English language learners are learning school subjects in English as part of the process in which they are learning the language itself. There are, however, reports that some schools are maintaining their Spanish-language bilingual programs by circumventing the new law in some manner. Nancy Ichinaga goes to the heart of the matter:

> California is having major problems enforcing Proposition 227 in many school districts with large Hispanic populations. As long as schools get federal money for having bilingual programs, as long as bilingual teachers are paid more than others, as long as there is a huge bilingual bureaucracy in the state and in the districts, there will be great resistance to giving up bilingual programs. It is too lucrative a jobs program for people to relinquish, even if it is being carried on the backs of children they profess to be for. (Ichinaga and Schieldge, p. 2)

Coincidental with the writing of this paper, on February 26, 2000, the *Los Angeles Times* reported that Nancy Ichinaga has been appointed by Governor Gray Davis to fill an opening on the California State Board of Education, the chief educational policymaking body in the state. This is good news, not only for Mrs. Ichinaga but also for those who believe as she does that English language literacy is the key to equal educational opportunity for our immigrant students. This prestigious appointment, in addition to the Salvatori Prize for American Citizenship, accords due recognition to an educator of genuine courage and steady commitment to high standards for all children.

References

Carter, Samuel C. (1999). *No excuses: Seven principals of low-income schools who set the standard for high achievement.* Washington, D.C.: The Heritage Foundation.

Fuetsch, Michelle. (1992, February 3). Two Inglewood schools defy odds, achieve excellence. *Los Angeles Times.*

Ichinaga, Nancy and Schieldge, Sandra. (1999). Battle against California's bilingual education establishment. Unpublished paper, Inglewood, Calif.

Business man, principal named to education panel. (2000, February 26). *Los Angeles Times.*

Bilingual Students and the MCAS

Some Bright Spots in the Gloom

Ralph E. Beals, Ph.D.
Amherst College

Rosalie Pedalino Porter, Ed.D.
READ Institute

Summary

The purpose of this study is to survey the participation and performance of Limited-English Proficient (LEP) students (often referred to as "bilingual" students) on the Massachusetts Comprehensive Assessment System (MCAS) in 1999. This study compares LEP students to each other by district, reporting on the rates of participation and the levels of achievement; identifies the districts where LEP students are achieving the highest passing scores on the MCAS for this cohort; and provides demographic data on the LEP students. This study focuses on the fourth-grade assessments in English language arts, mathematics, and science and technology, and covers all the districts (33) in which 10 or more LEP students were tested in one or more of these subjects, accounting for over 90 percent of the fourth-grade LEP students who were tested in the state. It is the first step in an ongoing study expected to continue and expand over the next several years.

A main conclusion of this survey is that the data collection and reporting by the Massachusetts Department of Education is seriously flawed, making it very difficult to interpret the results of the MCAS tests. These

are the major problems:

(1) data reported by the Department of Education in November 1999 are contradictory and inconsistent in regard to the numbers of LEP students tested—accurate figures were determined by extended investigation and educated estimates;

(2) LEP students who were eligible to take all MCAS tests in 1999 either did not take the math and science tests in half the districts surveyed or else their scores were not recorded (Table 1);

(3) for LEP students in grades 4, 8 and 10 who have not been in U.S. schools three years or longer and who are literate in Spanish, the Mathematics and Science tests could be taken in a bilingual (Spanish/ English) version of the test, ***but the Department did not mark the forms to identify who took the test in English or in the Spanish/ English version.***

Background

Some 45,000 students currently in Massachusetts classrooms entered the schools as Limited-English Proficient (LEP) children, often referred to as "bilingual students," or, in the newest usage, English language learners. For the sake of brevity and consistency, the term LEP will be used. These children started school without sufficient fluency and literacy in the English language to participate in regular classroom work in English. For this group, the Transitional Bilingual Education law, Chapter 71-A of the Massachusetts General Laws, was legislated in 1971 to give special help in the learning of English and in the learning of school subjects.

Under the guidelines of MCAS, LEP students are required to take the exams given in fourth, eighth, and 10th grade, in English, if they have been enrolled in U.S. schools for three years or longer (Memorandum of Commissioner D. Driscoll, December 8, 1998). As mentioned earlier, a Spanish bilingual version of the math and science tests is provided for those Spanish bilingual LEP students who have been in U.S. schools fewer than three years. Given this special accommodation for Spanish speakers, it is critical to know if students who take the bilingual version of the test demonstrate greater math and science proficiency, as a group, than those Spanish speakers who take the tests only in English. Without evidence that native language tests in math and science are of benefit to most students who use them, it seems hasty and perhaps even wasteful for the Department of Education to go forward with its translations of the MCAS subject matter tests into several other languages. Spanish speaking LEP students constitute approximately 70 percent of the LEP students in Massachusetts. There are few, if any, bilingual programs in the state that actually provide extended literacy and subject matter instruction in the native language, even in the major language groups, i.e., Vietnamese, Portuguese, Chinese dialects, and Haitian Creole. A wiser course may be for the Spanish bilingual tests to remain in use for several years so that data may be collected and the efficacy of the bilingual tests may be assessed before extending this approach to other language groups.

Until MCAS began the uniform assessment of all students in 1998, no reliable study had been conducted or published to document the academic achievement of LEP students in Massachusetts, or the particular districts that may be having greater or lesser success, or of the promising practices or programs being provided in the more successful schools. With the publication of the second year of test results in December 1999, it is now possible to present some descriptive data on LEP achievement in Massachusetts districts.

Scope and Limitations of the Study

The findings of this study, although narrowed to one grade level, do cover a substantial portion of the LEP students tested (over 90 percent) and provide an essential first step in an ongoing study that is expected to continue and expand for the next several years. The most complete MCAS results are reported for fourth-grade LEP students, and these are the focus of this study, along with 1998-1999 enrollment data and demographic data, i.e., length of time LEP students have been in U.S. schools, socioeconomic status of LEP students (percentage on free/reduced lunch). Comparisons are made of LEP student performance between districts (all 33 districts reporting 10 or more LEP students tested) and within districts (LEP students and their fully English-proficient classmates).

It is not within the scope of this study to address other issues at this time, such as (1) what criteria are used in each district to identify students as LEP; (2) what particular approach is used in each district to help these students, i.e., Transitional Bilingual Education (native-language instruction for several years) or mostly English as a Second Language emphasis (ESL), or two-way bilingual instruction; (3) comparisons between ethnic/language groups. These are areas in which the Department of Education collects data from the districts, and they will be the subject of future studies. At this time, "process" is not the focus but "outcomes" are. Once having established which districts have higher LEP test scores, it will then be useful to observe and record promising programs/practices in those districts and distribute the information statewide.

Slight Improvements, 1998–1999

The question of how many students take a test out of the total number eligible is crucial to the computation of student achievement and essential in making meaningful comparisons between districts. Although a direct comparison cannot be made between the rates of participation and performance between 1998 and 1999 due to a lack of reliable data, a well-educated estimate based on reasonable assumptions serves the purpose.

In 1998, 51.7 percent of the 2,891 LEP students enrolled in Massachusetts fourth-grade classrooms were reported as being "three years or longer in U.S. schools" and therefore required to take the MCAS in English. In fact, 66 percent of all LEPs enrolled took the English language arts test; and 83 percent took the math and science tests. In 1999, the Department of Education erroneously reported only 2,172 LEP students enrolled in Massachusetts fourth-grade classrooms when there were actually, by state census figures, 3,259. Of that number, 2,267 took the MCAS English language arts test at grade 4 (70 percent), 1,236 (38 percent) took the math test, and 1,188 (36 percent) took the science test (Tables 5, 6, 7). Since complete demographic data were not yet obtainable from the Department, we shall estimate that in 1999 approximately the same percentage of LEP fourth-graders were in their fourth year in U.S. schools (about 52 percent) as in 1998, and, therefore, required to take the MCAS. Using this reasonable estimate, since these proportions would not normally change very much from one year to the next, we find that in both years a greater number of LEP students participated in the MCAS English language arts test at the fourth-grade level than was required.

Test Score Index

The problem in 1999 is the low number of reported test scores in math and science, either because test booklets were incorrectly marked or because students did not take the required tests. Because there appears to be an undercounting of math and science tests, we used the figures for participation and passing rates on the 1999 fourth-grade English language arts test to rate LEP participation and performance in the 32 Massachusetts districts for which test results have been reported (Test Score Index, Table 3). The Test Score Index combines passing rates and participation in assessing the performance of districts. For example, Haverhill and Brookline had 22 and 21 LEP students respectively who took the fourth-grade English language arts MCAS. However, Haverhill tested 88 percent of the total enrolled while Brookline recorded only 40.3 percent of the students participating. In this case, even though only 50 percent of Haverhill's LEPs passed the test compared to Brookline's 86 percent, Haverhill is rated higher on the Index with .44 while Brookline scores .35.

The first column on the Test Score Index is an accurate figure, derived from the statewide school census, Table 5, on LEP students enrolled in each district by October of each school year. The second column, "LEP Students Tested," is taken directly from the *Summary of District Performance*, as is the percent passing reported in column 4. These figures are the most accurate we have found, with one caveat. Since the bilingual community is more mobile than the general population, with a high number of families moving in and out of districts every year, the enrollment fluctuates somewhat from the beginning to the end of each school year. A report by the U. S. General Accounting Office in Washington, D.C., assessing the educational challenges facing schools with LEP students included among its conclusions the fact that high levels of both family transiency and poverty exist in this population, both of which factors negatively affect children's academic development (cited in Porter, 1995, p. 12).

In instances where a larger number of LEP students were tested than appear to have been enrolled, it may be due to a number of factors. It may be that fourth-graders who were recently exited from special LEP programs were tested as LEP but no longer appear in the LEP enrollment statistics, new LEP fourth-graders arrived who were capable of taking the MCAS in English, or some other situation. Since this "overcount" only occurs in five of the 32 districts surveyed, it should not cause problems in the overall analysis.

Socioeconomic Status and Passing Scores

In the demographic description of LEP students included on the 1998 MCAS analysis published by the LAB for the Department of Education, the percentage of LEP students in a district who are on a free or reduce-price lunch program is used as an indicator of the socioeconomic status (SES) of such students for the district. It is a reasonable assumption that the percent of LEP students in the free/reduced lunch category would not have changed much in one year. In 1998, of all the 2056 fourth-grade LEP students who took the math test, 85.8 percent are reported to be eligible for free or reduced lunch. It is essential to bear in mind that low SES is one of the factors in academic underachievement, along with high family mobility and few years of formal education for parents. These factors are known to have a negative impact on a student's potential for school success, whether that child is bilingual or not.

Table 4, LEP Students Passing MCAS and Socioeconomic Status, provides—for those districts that gave the English language arts tests to a reportable number (10 or more) of LEP students—a view of the relationship between the percent passing the test, and the percent of LEP students who were on free or reduced-price lunch programs. For those few districts with especially low percentages of students on free/reduced lunch, the percentages of students passing the test is notably high, e.g., Newton, Arlington. Similarly, Holyoke, Lawrence and Lowell, with very high percentages of free/reduced lunch students, have the lowest fractions of passing scores.

Despite the overall strong relationship, there are a few exceptions to the rule: Consider Cambridge and Quincy, with high passing rates of 90 percent notwithstanding poverty rates of 80 percent. These and other exceptions will be watched closely in our annual reviews of the data.

Comparisons Within Districts

In order to avoid at least some of the socioeconomic disparities between districts—for example, comparing students in Fall River with students in Brookline—it is more useful to compare the differences in average scaled scores between LEP students and their English-speaking peers in the same districts. Some unexpected results emerge. Most noteworthy is the unique achievement of Chelsea, *the only district in the state in which the LEP students outscored "regular" students in any portion of the test, in particular, on the fourth-grade math test, 230 to 228.* Chelsea has introduced a new math program in the elementary schools, and this would seem to be a district where classroom observations should be considered. Some highlights of the wide range of differences in test scores within districts on the fourth-grade MCAS (Tables 5, 6, 7):

- English language arts: Five large districts' LEP students scored within 3-6 points of their non-LEP peers: Quincy, Boston, Cambridge, Fall River and New Bedford, while at the other end of the scale—Framingham, Methuen, Worcester and Amherst bilingual students scored 13-14 points below their classmates in the same district.
- In math and science, there are far fewer districts reporting scores, but there is a much wider gap in test scores. In math, Quincy, Boston, Fall River and New Bedford came within 1-6 points of their district averages while Framingham, Methuen and Waltham averaged 23-24 points lower; in science the results are similar, with Chelsea, Quincy, Brockton, Boston, New Bedford and Salem coming within 6-9 points of their districts, while Framingham, Methuen, Waltham and Haverhill are 22-27 points less.

District-by-District Performance

Due to the scarcity of data at the eighth- and 10th-grade levels, this study is restricted to the fourth-grade scores alone. In Tables 5, 6, and 7, we have compiled a district-by-district summary for English language arts, mathematics, and science and technology, for all districts enrolling sufficient numbers of bilingual children to be distinguished in the data set. The following information is reported on these tables:

- a rank ordered listing of districts with the largest number of LEP students participating in the grade 4 tests (32 districts, accounting for at least 93 percent of all the LEP students tested at grade 4) with the average scaled score for the district,

- the number of regular students (excluding students with disabilities) and their average scaled score, by district, for comparison, as well as the ratio of LEPs to regular students among fourth-graders in the district,
- a rank ordered listing of districts by LEP student performance as demonstrated by average scaled score.

Leaving aside the fact that LEP students as a whole scored lower than the native English-speaking students, useful information is obtained when comparing the performance of LEP students across districts, to discern where better learning is taking place and to arrive at some preliminary conclusions, as follows:

1. Again, the major caution: The greatest problem in determining accountability for LEP students is the poor quality of the data reported. Did administrators and teachers fail to understand the MCAS guidelines? Are LEP students not identified as "LEP," and are their scores included in the general population of "regular" students? Are LEP students simply not taking some of the tests they are required to take? For example, how can the Boston Public Schools be allowed to report that 639 LEP fourth-graders took the English language arts test but only 289 were tested in math and 272 in science? If a student is eligible to take the English test, that student is also required to take the other subject matter tests in English. Furthermore, as stated earlier, fourth-grade LEP students with fewer than three years attendance in U.S. schools may take the Spanish/English version of these two tests, which should result in even more test scores reported in math and science than in English language arts.
2. Quincy is the number one district in the state for high performing LEP students, since they are in fourth place in English and in first place in both math and science. Salem ranks second in LEP student performance: the district ranks second in science, fourth in math and 10th in English (Tables 5, 6, 7).
3. Chelsea scored second highest in math, third highest in science, while placing 23rd in the English test. For a district that has one of the lowest economic situations in the state, Chelsea shows outstanding achievement in outperforming wealthier districts.
4. While 22 of the 32 districts achieved the passing average scaled score of 220 in English, only six districts in math and eight in science averaged a passing score by LEP students. The data examined on Tables 5, 6, and 7, cover 93 percent of the LEP students in the state who were tested; the small remaining number are mainly in districts with too few LEP students to be reported. Table 3 establishes a hierarchy of district achievement by calculating a Test Score Index based on the proportion of stu-

dents tested in each district and the percentage of passing grades.

5. As stated earlier, it is not a part of this study to report any data on exactly how students are identified as LEP or on the exact kind of special instruction LEP students are being provided. Reliable information is not available on what goes on in each district or, in fact, in each classroom. No one knows whether or how much teaching is actually done in different languages, or what native language resources are available or how fully classroom instruction is aligned with MCAS frameworks.

Further Research/Further Considerations

Next to be addressed, in this ongoing analysis of the achievement of English Language Learners in Massachusetts public schools, are at least the following tasks, and probably more:

(1) analyses of the 2000 MCAS participation and performance by LEP students at all three grade levels, with the expectation that more complete data will be available;

(2) continued identification of districts with highest levels of academic performance by LEP students and highest levels of test participation; and

(3) a qualitative study of classroom observations in schools with a record of higher LEP student performance to identify promising instructional practices. With a third year's data to examine, it will be possible to review achievement in all districts with sufficient LEP students to detect trends in improvement or the lack thereof. If the promised, fuller data are indeed reported, it will be possible then to produce a more accurate and reliable picture than is possible at this time.

A qualitative study has been strongly recommended by the Massachusetts Bilingual Education Advisory Council and, if approved, is to be carried out during the 2000-01 school year under the new research arm of the Department of Education. Once having identified the particular schools (six to 10 of them) where better achievement for LEP students is documented, a team of researchers will visit classrooms to record the promising practices observed. What are teachers doing and in what language and with what materials to produce better learning? A report linking good teaching with documented student success will provide urgently needed information to all districts in Massachusetts striving to improve educational opportunities for their LEP students. This is a welcome development in bilingual education research and reflects the current trend across the country in this field. After focusing for three decades on "process" only and avoiding the important matter of "outcomes" in student achievement, Massachusetts can make a valuable contribution to the research literature with this study.

In an interview on July 7, 2000, with Jeff Nellhaus, associate commissioner in charge of assessment, Porter was informed that the next MCAS report will make clear which LEP students have taken the subject matter tests in the Spanish/English version. We trust the performance of LEP students who take the test in the English version will be reported separately from those who take the bilingual version. It is not possible to gauge the benefits of Spanish language assessments (and of the future use and value of MCAS tests in other languages) without accurate reporting on this point.

Conclusion

We have reported what can be determined to date about the participation and performance of LEP students in two years of application of the MCAS. Given the wide diversity of the population of LEP student backgrounds in language, ethnicity, and earlier education in other countries, much more data are needed. When accurate information is compiled and available to researchers in a timely fashion—on LEP student SES, mobility, years enrolled in special programs, attendance rates and dropout rates—these data, coupled with MCAS scores, will provide a much more realistic understanding of expectations for this population. MCAS test scores alone cannot present the entire account but they do provide the only fair, objective, neutral measure of how these students are meeting the standards set by the commonwealth for academic achievement.

This ongoing study should be considered a pioneering effort since almost no research on Massachusetts LEP student achievement has yet been published. Some basic questions about Transitional Bilingual Education programs are raised by the MCAS data. The theoretical basis for this education model is that LEP students will learn their school subjects more effectively if they are taught in their first (or native, primary, home) language than if they are taught subject matter in the second language (English). Yet LEP students are scoring higher on the English language arts test than on math and science in almost every district (Table 1.) Are so-called bilingual programs actually doing most of the teaching in English? Is the quality of the English language teaching superior to the quality of the math and science teaching in most bilingual classrooms? Are some of the "bilingual" programs really English immersion programs in disguise? Is the quality of Spanish-language instruction less adequate, or is there a lack of Spanish language textbooks in math and science? These issues will be addressed when more information is known.

Education reform efforts begun in different states since the 1970s are beginning to bear fruit. A Rand Corporation study released on July 26, 2000, announces the welcome news that math scores are rising across the country, showing more progress in this decade than in the previous 20 years, based on testing conducted by the National Assessment of Educational Progress (NAEP). The study finds that "education reforms in the late 1980s and early 1990s have paid off in terms of higher math scores for public school students, especially among black and Hispanic students," and attributes these gains principally to "...state-sponsored pre-kindergarten programs, targeting more resources for schools in lower-income areas, and *using test scores to highlight differences in performance between schools.* [author's emphasis]" (Fialka, p. A28).

Acknowledgments

Jodie Zalk, Carolyn Faria, and Lisa Vinikoor at the Massachusetts Department of Education gave valuable assistance to this project by providing the statistical data we requested promptly and in a very helpful manner. Professor Christine Rossell at the Boston University Political Science Department rendered invaluable advice and contributed her time and expertise in two reviews of this study.

References

Driscoll, D. (1998, December 8). Memorandum to Massachusetts Board of Education. Malden, Mass.

Fialka, J. (2000, July 26). Math scores rise in U.S. public schools with minorities and Texas leading way. Washington, D.C.: *The Wall Street Journal.*

Lachat, M. A., Holt, K., and Beaudet, S. (1999, September). *The participation and performance of limited-English proficient students in the 1998 Massachusetts Comprehensive Assessment System.* Providence, R.I.: Northeast and Islands Regional Educational Laboratory at Brown University (LAB).

Massachusetts Department of Education. (1999, November). *The Massachusetts Comprehensive Assessment System: Summary of district performance.* Malden, Mass.

Massachusetts Department of Education. (1999, November). *Massachusetts Comprehensive Assessment System: Report of 1999 state results.* Malden, Mass.

Porter, R. P. (1995, Spring). A review of the U.S. GAO study on limited-English students. Washington, D.C.: *READ Perspectives,* II-1, 9-23.

Table 1.
Number of LEP Students Participating in MCAS Percent of LEP Students Passing MCAS 1998 and 1999

	1998		1999[1]	
	Number Tested	**% Passing**	**Number Tested**	**% Passing**
Grade 4–				
English Language Arts	1,908	49.1	2,267	56
Math	2,395	34.7	1,236	40
Science & Technology	2,390	50.9	1,188	52
Grade 8–				
English Language Arts	758	47.6	780	52
Math	1,042	20.0	525	12
Science & Technology	1,044	13.1	535	9
Grade 10–				
English Language Arts	717	36.0	647	31
Math	1,051	18.7	505	5
Science & Technology	1,055	21.7	491	14

- The 1998 data were prepared for the Massachusetts Department of Education by the LAB at Brown University, August 1999. The 1999 figures were published by the Massachusetts Department of Education in their 1999 *Report of State Results, and Summary of District Results,* both November 1999.
- There appears to be a curious inversion between 1998 and 1999 in the proportions of LEP students tested in each subject. In 1998, far more students were tested in Math and Science than in English Language Arts; in 1999 the opposite is reported.

---

[1] 1999 data from *The Massachusetts Comprehensive Assessment System: Summary of District Results.* Massachusetts Department of Education, November 1999.

Table 2.
Statewide Comparison of LEP to All Students Average Scaled Scores—1999 MCAS Performance

Percentage of LEP Students at each Performance Level:

	Advanced	Proficient	Needs Improvement	Failing ■		Average Score—LEP Students	Average Score—All Students
				(Tested)	(Absent)		
Grade 4—							
English	0	3	53	43	0	222	231
Math	1	5	34	61	0	218	235
Science & Technology	0	7	45	48	0	220	240
Grade 8—							
English	0	14	39	47	1	221	238
Math	1	3	8	87	0	207	226
Science & Technology	0	2	7	91	0	204	224
History & Social Studies	0	0	9	91	0	206	221
Grade 10—							
English	0	6	25	66	3	213	229
Math	0	1	4	92	4	203	222
Science & Technology	0	1	13	80	6	208	226

■ "Failing" is reported in two columns. According to Massachusetts Department of Education guidelines, a scaled score below 220 is reported as "Failing-Tested." Students who were absent without a medically documented excuse from any subject area MCAS test were assigned the minimum scaled score of 200 and a performance of "Failing-Absent" for that subject area

Table 3.
Test Score Index*
1999 Grade 4 MCAS—LEP Students Participation English Language Arts

District	LEPs Enrolled 1998-1999**	LEPs Tested 1999	Fraction Tested ×	% Passing =	Test Score Index
Boston	972	639	.657	66	43
Lowell	313	165	.527	22	12
Lawrence	318	150	.471	31	15
Holyoke	184	132	.717	31	22
Springfield	257	124	.482	38	18
Quincy	19	86	1.000	90	90
Chelsea	79	85	1.000	52	52
Lynn	120	84	.700	46	32
Fitchburg	83	71	.855	45	38
Worcester	145	60	.413	37	15
Framingham	107	50	.467	68	32
Salem	81	50	.617	82	51
New Bedford	51	45	.882	58	51
Newton	44	29	.659	96	63
Brockton	72	28	.388	50	19
Methuen	45	27	.600	44	26
Randolph	38	27	.710	85	60
Revere	35	24	.685	50	34
Cambridge	47	23	.489	91	45
Fall River	25	23	.920	78	72
Haverhill	25	22	.880	50	44
Brookline	52	21	.403	86	35
Somerville	8	21	1.000	81	81
Chicopee	39	17	.436	41	18
Arlington	16	15	.937	100	94
Westfield	31	13	.419	62	26
Amherst	21	12	.571	100	57
Taunton	21	12	.571	42	24
Woburn	5	11	1.000	100	100
Leominster	7	10	1.000	100	100
Malden	24	10	.416	70	29
Marlborough	13	10	.769	80	62

*Test Score Index concept recommended by Dr. C. Rossell, Boston University, Department of Political Science.

** Enrollment figures provided by Technology Office, Massachusetts Department of Education, School Census Report, Table 5.

Table 4.
LEP Students Passing MCAS and Socioeconomic Status

1999 MCAS English Language Arts Test Scores—Fourth-Grade LEP Students
Percentage of LEP Fourth-Graders on Free/Reduced Lunch

District	No. LEPs Tested	Percent LEPs Passing 4th-Grade MCAS—English	Percent LEPs on Free/Reduced Lunch*
Arlington	15	100	27
Amherst	12	100	27
Leominster	7	100	26
Woburn	5	100	16
Newton	29	96	30
Cambridge	23	91	79
Quincy	86	90	81
Brookline	21	86	42
Randolph	27	85	29
Salem	50	82	88
Somerville	21	81	47
Marlborough	10	80	75
Fall River	25	78	83
Malden	10	70	88
Framingham	50	68	91
Boston	639	66	88
Westfield	13	62	23
New Bedford	45	58	92
Chelsea	85	52	93
Brockton	28	50	91
Revere	24	50	83
Haverhill	22	50	29
Lynn	84	46	84
Fitchburg	71	45	94
Methuen	27	44	23
Taunton	12	42	67
Chicopee	17	41	100
Springfield	124	38	92
Worcester	60	37	96
Lawrence	150	31	99
Holyoke	132	31	100
Lowell	165	22	97

* Figures obtained from LAB report prepared for Massachusetts Department of Education, 1998 school year, Table D4, pp. 33-36, and from Technology Office, Massachusetts Department of Education.

Table 5.
Grade 4—English—1999

This listing covers all districts that tested 10 or more LEP students.

ID	District	LEP Students No.	LEP Students Avg. Score	Regular Students No.	Regular Students Avg. Score	Ratio No.'s LEP to Reg (%)	Districts Ranked by LEP Score Avg. Score	Districts Ranked by LEP Score District
035	Boston	639	223	3039	226	21	233	Woburn
160	Lowell	165	215	994	226	17	232	Newton
149	Lawrence	150	217	727	224	21	232	Arlington
137	Holyoke	132	216	303	228	44	231	Quincy
281	Springfield	124	217	1348	229	9	229	Cambridge
243	Quincy	86	231	474	234	18	229	Brookline
057	Chelsea	85	219	296	225	29	227	Randolph
163	Lynn	84	218	881	227	10	226	Westfield
097	Fitchburg	71	219	376	230	19	226	Malden
348	Worcester	60	218	1453	232	4	225	Salem
100	Framingham	50	222	513	236	10	224	Somerville
258	Salem	50	225	322	231	16	223	Boston
201	New Bedford	45	222	953	227	5	223	Fall River
207	Newton	29	232	682	242	4	223	Amherst
044	Brockton	28	220	1089	228	3	223	Leominster
181	Methuen	27	220	440	233	6	223	Marlborough
244	Randolph	27	227	275	233	10	222	Framingham
248	Revere	24	221	380	233	6	222	New Bedford
049	Cambridge	23	229	394	232	6	221	Revere
095	Fall River	23	223	822	228	3	220	Brockton
128	Haverhill	22	218	573	230	4	220	Methuen
046	Brookline	21	229	355	239	6	220	Taunton
274	Somerville	21	224	284	231	7	219	Chelsea
061	Chicopee	17	218	394	230	4	219	Fitchburg
010	Arlington	15	232	282	240	5	218	Lynn
325	Westfield	13	226	368	233	4	218	Worcester
008	Amherst	12	223	178	236	7	218	Haverhill
293	Taunton	12	220	551	231	2	218	Chicopee
347	Woburn	11	233	316	239	3	217	Lawrence
153	Leominster	10	223	410	234	2	217	Springfield
165	Malden	10	226	332	232	3	216	Holyoke
170	Marlborough	10	223	286	234	3	215	Lowell
308	Waltham	8	–	291	234	3		
TOTALS		2,096		20,381		10.3%		
AVERAGE			222.8		231.7			
STATEWIDE		2,267	222	60,348	234	3.8%		
Coverage		93%		34%				

Table 6.
Grade 4—Math

This listing covers all districts that tested 10 or more LEP students.

ID	District	LEP Students No.	LEP Students Avg. Score	Regular Students No.	Regular Students Avg. Score	Ratio No.'s LEP to Reg (%)
035	Boston	289	220	3684	226	8
149	Lawrence	176	212	790	223	22
160	Lowell	131	213	1072	227	12
137	Holyoke	116	215	362	227	32
281	Springfield	65	218	1518	227	4
348	Worcester	64	216	1558	234	4
097	Fitchburg	43	213	419	228	10
243	Quincy	43	233	532	239	8
100	Framingham	41	218	548	241	7
057	Chelsea	37	230	333	228	11
163	Lynn	28	210	988	226	3
258	Salem	27	224	363	234	7
201	New Bedford	19	220	1012	225	1.9
181	Methuen	16	215	458	239	3
308	Waltham	12	212	312	235	4
095	Fall River	11	227	847	228	1.3
128	Haverhill	11	213	618	229	1.8
044	Brockton	10	219	1129	228	0.9
274	Somerville	9	-	309	234	3
293	Taunton	9	-	572	231	1.6
244	Randolph	7	-	299	231	2.3
207	Newton	5	-	724	253	0.7
170	Marlborough	4	-	303	238	1.3
248	Revere	4	-	415	232	1.0
153	Leominster	3	-	423	238	0.7
046	Brookline	2	-	382	245	0.5
049	Cambridge	2	-	418	235	0.5
165	Malden	2	-	349	234	0.6
061	Chicopee	1	-	430	230	0.2
008	Amherst	0	-	203	243	0
010	Arlington	0	-	310	246	0
325	Westfield	0	-	396	236	0
347	Woburn	0	-	330	244	0
TOTALS		**1,187**		**22,406**		**5.3%**
AVERAGE			**218.2**		**233.8**	
STATEWIDE		**1,236**	**218**	**63,590**	**237**	**1.9%**
Coverage		**96%**		**35%**		

Districts Ranked by LEP Score Avg. Score	District
233	Quincy
230	Chelsea
227	Fall River
224	Salem
220	Boston
220	New Bedford
219	Brockton
218	Springfield
218	Framingham
216	Worcester
215	Holyoke
215	Methuen
213	Lowell
213	Fitchburg
213	Haverhill
212	Lawrence
212	Waltham
210	Lynn

Table 7. Grade 4—Science—1999

This listing covers all districts that tested 10 or more LEP students.

ID	District	LEP Students No.	LEP Students Avg. Score	Regular Students No.	Regular Students Avg. Score	Ratio LEP to Reg (%)	Districts Ranked by LEP Score Avg. Score	Districts Ranked by LEP Score District
035	Boston	272	220	3678	229	7	236	Quincy
149	Lawrence	152	214	810	226.	19	229	Salem
160	Lowell	128	217	1079	231	12	225	Chelsea
137	Holyoke	113	217	360	232	31	225	Brockton
348	Worcester	64	219	1558	238	4.1	224	Framingham
281	Springfield	61	218	1507	233	4	222	New Bedford
100	Framingham	52	224	543	247	10	220	Boston
243	Quincy	46	236	530	243	9	219	Worcester
097	Fitchburg	43	218	417	236	10	218	Springfield
057	Chelsea	35	225	338	231	10	218	Fitchburg
163	Lynn	28	215	998	230	2.8	218	Methuen
258	Salem	27	229	361	238	7	217	Lowell
201	New Bedford	20	222	1008	231	2.0	217	Holyoke
181	Methuen	16	218	460	240	3	215	Lynn
128	Haverhill	15	214	618	237	2.4	214	Lawrence
308	Waltham	12	213	311	240	4	214	Haverhill
044	Brockton	10	225	1135	233	0.9	213	Waltham
274	Somerville	9	-	308	238	2.9		
293	Taunton	9	-	565	238	1.6		
095	Fall River	8	-	850	236	0.9		
244	Randolph	6	-	303	237	2.0		
170	Marlborough	5	-	301	242	1.7		
207	Newton	5	-	728	251	0.7		
248	Revere	5	-	416	236	1.2		
049	Cambridge	4	-	419	237	1.0		
046	Brookline	2	-	382	248	0.5		
153	Leominster	2	-	426	241	0.5		
165	Malden	2	-	352	238	0.6		
061	Chicopee	1	-	433	236	0.2		
008	Amherst	0	-	201	244	0		
010	Arlington	0	-	306	250	0		
325	Westfield	0	-	395	242	0		
347	Woburn	0	-	328	249	0		
TOTALS		**1,152**		**21,395**		**5.4%**		
AVERAGE			**220.7**		**238.8**			
STATEWIDE		**1,188**	**220**	**63,688**	**242**	**1.9%**		
Coverage		**97%**		**34%**				

1999 data from The Massachusetts Comprehensive Assessment System: Summary of District Results. Massachusetts Department of Education, November 1999.

Different Questions, Different Answers

A Critique of the Hakuta, Butler, and Witt Report, "How Long Does It Take English Learners To Attain Proficiency?"

Christine H. Rossell, Ph.D.

Kenji Hakuta, Yuko Goto Butler, and Daria Witt begin their paper[1] with the statement:

> One of the most commonly asked questions about the education of language minority students is how long they need special education services, such as English as a Second-Language (ESL) and bilingual education (p.1).

Unfortunately, they do not present any research on this issue in their paper. Nevertheless, this does not stop them from concluding:

> The data would suggest that policies that assume rapid acquisition of English—the extreme case being Proposition 227 that explicitly calls for "sheltered English immersion during a temporary transition period not normally intended to exceed one year"—are wildly unrealistic (p. 13).

Although they appear not to know it, there is no research presented in this paper that tells us how long limited-English proficient (LEP) students should be in a sheltered English immersion classroom. The research that is presented is on a different issue: how long it takes a limited-English proficient student, on average, to attain the average English language achievement of fluent English speakers or a test publisher's criterion for English proficiency.

The authors are simply wrong in believing that knowing how long it takes an LEP child to achieve parity with native English speakers, or to be classified "proficient" on an English proficiency test, tells us how long they need special education services or how long they should be in a sheltered immersion classroom.

The Data

The Hakuta et al. study consists of LEP students in four samples, two of them in school districts in the San Francisco Bay area and two of them in Canada. They collected and analyzed the data in School Districts A and B in California themselves and reanalyzed summary data on the two Canadian samples that were reported in Wright and Ramsey, 1970; Cummins, 1981; and Klesmer, 1993.

School districts A and B in California vary considerably in socioeconomic status (SES). The sample of LEP students in district A consists of all 1,872 LEP students in Grades 1-6 in spring 1998 who had been in the district since kindergarten and were classified at that time as LEP. About half were Vietnamese speakers and half Spanish speakers. According to the authors, the district has been on a state waiver from bilingual education, and has never provided systematic instruction through the native language. The percentage of students on free or reduced-price lunch is low—35 percent—and their annual redesignation rates from LEP to English proficient are high, about four times the state average.

District B, by contrast, has a free or reduced lunch rate of 74 percent-twice that of District A. The sample in District B consists of 122 Spanish speakers in grades 1, 3, and 5 during the spring of 1998, randomly selected from the students who had been in the school district since kindergarten, were classified LEP at that time, and who attended high poverty schools. Some of these LEP students were in bilingual education and some in ESL, although the authors assert there was no difference in achievement between students in the two programs.

The Toronto data reported in Wright and Ramsey (1970) and Cummins (1981) consists of 1,200 immigrant children learning English as a second language selected from a survey of 25 percent of the Toronto school system's classrooms in Grades 5,7, and 9, who were of varying length of residence in Canada. Although the authors do not specify what language the students were instructed in, it was undoubtedly English since that is the normal approach in Canada to educating immigrant children.

The North York, Ontario data, reported in Klesmer (1993), consisted of a randomly selected sample of 285 ESL students and 43 native English-speaking students who were controls. All students were 12 years old and most of them were in the seventh grade, but their length of residence ranged from six months to almost six years. Since the students are called "ESL" students, we can assume they are being instructed in English.

The research design varies across studies. The data from Toronto and North York are cross-sectional. They consist of students at fixed grade levels who differ in their length of residence in Canada. The data from Districts A and B in California are longitudinal and consist of the more stable LEP students, those who had been in the school district since kindergarten and were classified LEP at that time. The Canadian data are not longitudinal, but they will be biased only if the composition of the students being studied changes over time in a way that influences the outcome. I am not aware of any such changes, and the authors do not mention any.

Hakuta, Butler and Witt's findings are divided into oral English and academic English, a distinction that is commonly made, but in fact is not based on research or experience. All English is academic English, and there really is no way to separate academic English from oral English. Moreover, there is extensive research, discussed below, that contradicts the notion that oral English proficiency is "non-academic" and only tests whether a child understands the English language. Therefore, I have changed their term "academic" to "written" to conform to what the tests actually assess and to maintain the useful distinction between oral and written tests.

Hakuta, Butler, and Witt's findings on how long it takes LEP students to attain English "proficiency" are summarized in Table 1. English proficiency was assessed in the two California school districts by means of a specific criterion on oral and written English proficiency tests and in the two Canadian samples by means of parity with native English speakers on oral and written standardized achievement tests. Their findings show that, on average, it takes anywhere from two years to perhaps forever to attain the criterion for English proficiency in the California school districts, and from nine years to perhaps forever to attain parity with English native speakers in the Canadian samples.

Table 1.

Number of Years It Takes LEP Students To Attain English Language Parity with Native English Speakers or a "Proficient" Score on an English Proficiency Test in the Hakuta, Butler, and Witt Report

	CALIFORNIA		CANADA	
	ENGLISH PROFICIENCY TESTS		STANDARDIZED ACHIEVEMENT TESTS	
	(criterion)		(parity)	
	DISTRICT A (higher SES)	DISTRICT B (lower SES)	TORONTO	NORTH YORK
ORAL	2 - 5 years		9 - 11 years	Not attained after 5 years
WRITTEN	4 - 7 years after 5 years	Not attained		Not attained after 5 years

English proficiency is negatively correlated with socioeconomic status. Table 1 shows that the students in District A, 35 percent of whom are on free lunch, achieve parity with English speakers before District B students, 74 percent of whom are on free lunch. There is also a correlation between SES and English proficiency within districts. Hakuta, Butler, and Witt separated the students in District A into the poverty levels of their school—10 percent, 25 percent, 50 percent, and 70 percent free lunch. They found that the higher the school poverty level, the lower the level of English "proficiency." In District B, they analyzed parents' self-reported formal education and found that the higher the parents' educational level, the higher the LEP students' test scores.

Hakuta, Butler, and Witt's findings are both believable and consistent with other research. Where I disagree with them is with regard to what these test results mean and the policy implications.

Unwarranted Conclusions

Hakuta, Butler, and Witt jump to the conclusion that the number of years it takes LEP students to reach the average for native English speakers or the publisher's criterion for English proficiency is the number of years they need special education services. There are two reasons why this conclusion is unwarranted. First, parity with English speakers on English proficiency tests or standardized achievement tests is a badly flawed standard for determining fluency in English. Half of all native English speakers cannot achieve the average standardized test score for native English speakers, and almost as large a percentage cannot achieve the publisher's criterion for English proficiency. If the students are of low SES, as is typically the case with immigrant children, more than half will not achieve the average for English speakers or the criterion for English proficiency, no matter how fluent they are in English. This failure to understand what test scores mean and their biases is probably one of the most common errors made by reporters, politicians, other laymen, and even by experts in the field, and it is disheartening to see the mistake made once again.

The second reason why one cannot jump to the conclusion that the number of years it takes LEP students to reach the average for native English speakers or the publisher's criterion for English proficiency is the number of years they need special education services is that the research design used by Hakuta et al., and the studies they analyze, do not allow us to draw such conclusions. To determine whether an LEP child is better off with special education services than without requires the following design. LEP children must be randomly assigned to a group that receives no special education services and to groups that receive some carefully documented special service over different periods of time. The achievement of students in these groups is then compared and a statistical analysis performed to determine if there is a significant difference between the groups.[2]

The research design that would definitively answer the question of how long LEP students should remain in a structured immersion classroom, a particular type of special education service, would randomly assign non-English speaking students in each grade to a mainstream classroom and to a structured immersion classroom. These students would be tested initially and then at monthly intervals. The point at which the students in the mainstream classroom outperform the students in the structured immersion classroom on the tests is the point at which a student is better off in the mainstream classroom than in a structured immersion classroom. If over several years, the students who entered the mainstream classroom sooner outperform the students who entered later, then a mainstream classroom is a better environment from the start regardless of the short-term data.

Although it might appear that the structured immersion classroom would be superior to the mainstream classroom for a very long time, that is probably not the case. My own estimate would be that sometime during the first year there is probably no difference between the structured immersion classroom and the mainstream classroom because although the structured immersion classroom may be a better environment in the beginning, it has the following negative characteristics: (a) it has a slower pace which will begin to negatively affect students who can understand English, and (b) it has no English-speaking role models. Students interact with other students whose English is also imperfect and this can become a problem because students emulate the English of their peers. If they cannot understand English, they are better off in a structured immersion classroom. But when they reach the point where they can understand English, they will speak like their classmates and they will be better off if their classmates are speaking grammatical English.

Hakuta et al. do not do this analysis, nor do they present research that has done this. Therefore, they cannot legitimately claim that their study tells us how long LEP children should receive special education services or be in a structured immersion classroom.

Norm-Referenced Tests

As noted above, the first reason why the findings of Hakuta et al. cannot tell us how long a child needs special education services is that the instruments and procedures used to measure English proficiency are flawed both in design and in use. Table 2 summarizes the tests and standards used in the samples analyzed by Hakuta, Butler, and Witt. The table is divided into the same categories as Table 1, but the cells now contain the type of test and the criterion used for English proficiency. In addition, I have added a row indicating the biases of the English proficiency tests used in the California school districts and the standardized achievement tests used in the Canadian samples in determining whether a student is fluent in English.

English proficiency tests are a type of norm-referenced test given to students identified by a home language survey as coming from a home where someone speaks, or has spoken, a language other than English. They are also given to students who have already been identified as LEP in order to determine if they are now English proficient. The state of California approves the following English proficiency tests which have both oral and written forms: the BINL, BSM I/II, Pre-IPT, IPT I/II, pre-LAS, LAS I/II, the Woodcock-Muñoz Language Survey, and the QSE.

Two of these tests are used in the California districts studied by Hakuta, Butler, and Witt. District A uses the IPT and the oral results are shown in Table 1. District B uses the Woodcock-Muñoz Language Survey. The results for the written portion are also in Table 1.

As shown in Table 2, the California school districts use a specific English proficiency criterion established by the publisher in the case of the IPT and the Woodock-Muñoz Language Battery and by the district in the case of the MacMillan Informal Reading Inventory.[3] The Canadian studies use parity with English speakers on oral and written standardized achievement tests.

Children can be completely fluent in English, indeed they can know no language other than English, and yet fail to achieve the publisher's criterion for English proficiency. All language proficiency tests, whether they are administered only to LEP students (and called English proficiency tests) or to English speaking students (and called achievement tests), are norm-referenced on fluent English speakers and are tests of the ability to speak and understand a language and tests of academic ability in that language. The publishers select a score on the English proficiency tests that they claim denotes whether a student is a fluent English speaker, but in fact there are English monolingual students who will score below whatever score they select unless it is zero. Typically the publishers select a score that can only be achieved by about 60 percent to 70 percent of the English monolingual students.

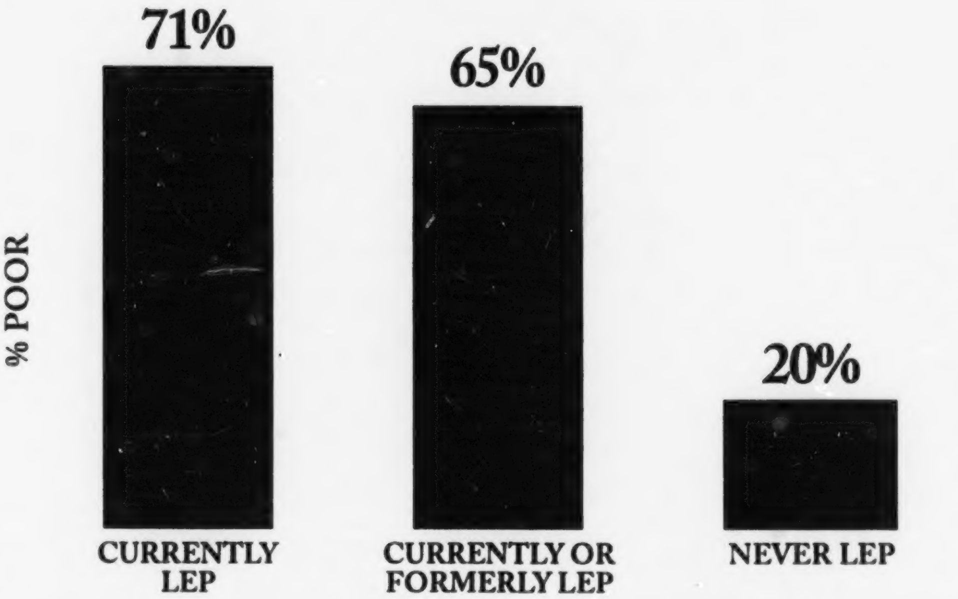
Table 2.
Tests and Standards Used in Hakuta, Butler, and Witt Report and Their Biases

	CALIFORNIA		CANADA	
	ENGLISH PROFICIENCY TESTS		STANDARDIZED ACHIEVEMENT TESTS	
	(criterion)		(parity)	
	DISTRICT A (higher SES)	DISTRICT B (lower SES)	TORONTO	NORTH YORK
ORAL TESTS	Publisher's Standard IPT English proficiency test		Parity with native speakers Picture Vocabulary Test and unspec. test of grammar	Parity with native speakers Unspecified test
WRITTEN TESTS	District Standard MacMillan Informal Reading Inventory and unspec. writing test	Publisher's Standard Woodcock-Muñoz Language Battery English proficiency test		Parity with native speakers Degrees of Reading Power
BIASES	1) PUBLISHER'S STANDARD CAN ONLY BE OBTAINED BY 60 PERCENT TO 70 PERCENT OF ENGLISH MONOLINGUAL STUDENTS 2) LOWER SOCIOECONOMIC (SES) STUDENTS SCORE LOWER THAN HIGHER SES STUDENTS EVEN IF ALL ARE FLUENT IN ENGLISH		LOWER SES STUDENTS SCORE LOWER THAN HIGHER SES STUDENTS EVEN IF ALL ARE FLUENT IN ENGLISH	

Using parity with native speakers on standardized achievement tests as a means of determining English fluency, as is done in the Canadian studies, is biased by the fact that standardized achievement tests rank order students and this rank ordering is highly correlated with socioeconomic status. The test scores do not tell us what students know. They only tell us who knows more and who knows fewer answers to the items on the test. These items are deliberately selected to produce a normal curve among English speaking students, and the test scores are highly correlated with socioeconomic status.

The analyses of the Canadian samples presented in Hakuta, Butler, and Witt are biased by the fact that immigrant children are of lower social class than non-immigrant children. This is shown in Figure 1 which presents data on the percentage of students on free or reduced lunch by LEP status in spring 1997 in a medium sized California school district of about 35,000 students. The percentage of currently LEP students who are poor is 71 percent and the percentage of currently or formerly LEP students who are poor is 65 percent. The latter group includes formerly LEP students so as to include as many immigrant students as possible, not just those who continue to score low on English proficiency and standardized achievement tests. LEP students have more than three times the percentage poor of non-LEP students.

Figure 1.
% Poor in California School District by LEP Status, Spring 1997



To understand how this affects the standardized achievement test results in the Canadian samples in Hakuta, Butler, and Witt, we need to look at the relationship between poverty and standardized test scores in an English speaking sample. Figure 2 shows a box and whiskers plot of the CAT5 achievement—vocabulary, reading comprehension, math analysis, and math computation—of all secondary students (including poor students) at the top of the page and the achievement of only the poor students under it, among students who are fluent English speaking and who have never been classified LEP. The black line across each box is the median[4] achievement for each group. The box itself is the interquartile range—the range from the 25th to the 75th percentile that contains 50 percent of the cases. The horizontal lines at each end of the vertical lines are the maximum scores.

I have added the average scores for each group of students below each subtest. Note the 36th percentile, the most common standard in California for redesignating an LEP student as English proficient.

The analysis in Figure 2 is similar to the analysis of the Canadian studies presented in Hakuta, Butler, and Witt. The studies they analyzed compared the standardized test scores of immigrant children to English native speakers. To show the bias produced by the fact that immigrant children are of lower social class than all children, I have compared poor English-speaking children to all English native speakers. English speaking poor children are not even close to attaining parity with all English speakers, although both groups are fluent English speakers. Indeed, this is exactly the problem with standardized, achievement tests—they merely rank order students on knowledge of the items on a test and they cannot tell the difference between students who do not know English and students who do not know the answer. Thus, any comparison of the achievement of a high-poverty group, such as immigrant children, to the achievement of all students, as Hakuta, Butler, and Witt have done, will find the poorer group performs worse.

English Proficiency Tests

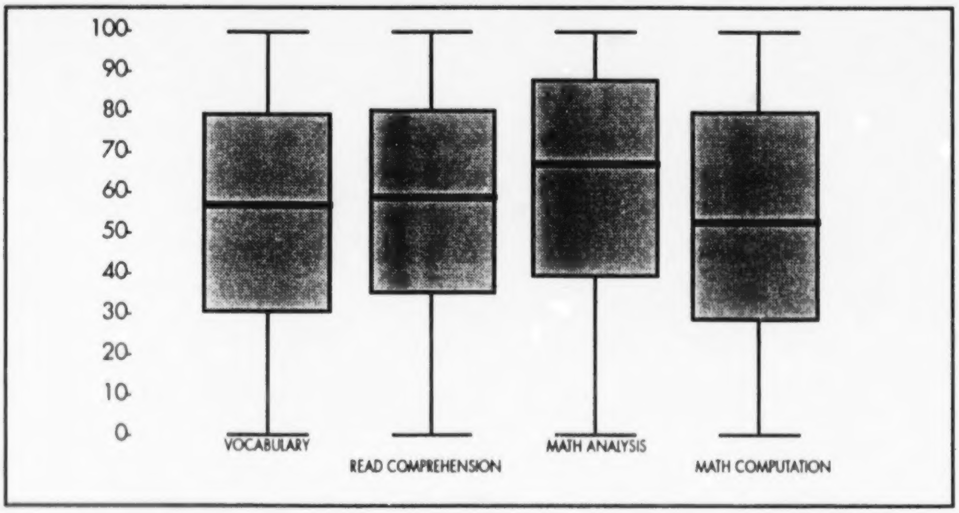
All English proficiency tests, whether oral or written, are known to be unreliable—that is, you cannot get the same outcome in subsequent tests of the same child—and invalid—that is, they do not accurately determine who is LEP (Baker and Rossell, 1987; Rossell and Baker, 1988). On the face of it, oral English proficiency tests would seem to be better than a written test at determining whether a child knows enough English to function in a mainstream classroom because the child doesn't have to know how to read or write to take an oral proficiency test.

Unfortunately, oral English proficiency tests are no better than written English proficiency and standardized achievement tests, and for many of the same reasons. Moreover, they have some additional problems that written proficiency tests do not have. In oral tests, students are asked questions that require that they not only know English, but understand and remember the question and have the self-confidence to stand up to a stranger when the question is not understood. Thus, contrary to the assertions of Hakuta, Butler, and Witt, oral tests are as "academic" as written tests. Like standardized achievement tests administered to the English speaking stu-

Figure 2.

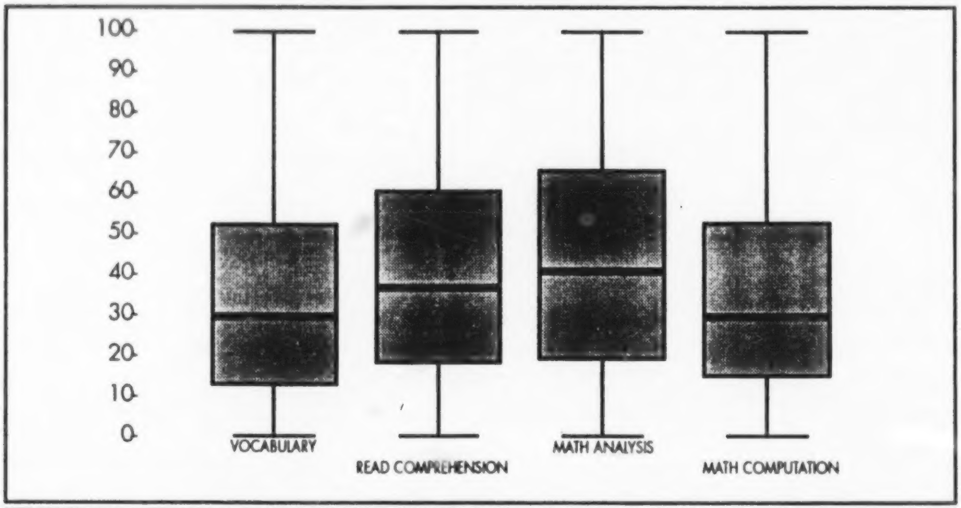
Achievement on CAT5 of All Fluent English-Speaking Secondary Students and Poor, Fluent English-Speaking Secondary Students in a California School District, Spring 1997

ALL STUDENTS



	Vocabulary	Reading Comprehension	Math Analysis	Math Computation
Average Score	54	56	61	53

POOR STUDENTS



	Vocabulary	Reading Comprehension	Math Analysis	Math Computation
Average Score	35	39	43	36

dent body, and written English proficiency tests administered only to LEP students, oral proficiency tests cannot tell the difference between a student who does not know English and a student who does not know the answer. They are normed on an English-speaking body and the same arbitrary cut-off points are used.

Despite these problems, language proficiency tests are used everywhere as a means of identifying whether a student is LEP and English proficient, and their use is codified in state legislation and court decisions. New York City, for example, uses the L.A.B. whose oral portion was normed in 1981-82 and whose written portion was normed in 1985, in both instances on an English-speaking citywide population. The criterion selected for determining whether a child is fluent English proficient in New York City is currently the 40th percentile on the L.A.B. In many California school districts, including the two studied by Hakuta, et al., the standard is the 36th percentile. It is a mathematical principle that 40 percent of the norming population scores at the 40th percentile, and 36 percent scores at the 36th percentile. If the L.A.B. were administered citywide in New York City, a minimum of 40 percent of the children in the city, almost all of whom are English native speakers, would fail to be classified as English proficient. If the tests used in the California districts were administered to all students, a minimum of 36 percent would fail to be classified as English proficient, even if the only language they know is English.

To the extent that these students are of lower SES, even higher percentages will fail to be classified as English proficient. If we look at the box on the bottom of Figure 2, the analysis of the achievement of poor, English speaking students, we can see that about 50 percent of these students would fail to be classified English proficient if the standard were the 36th percentile.

Interestingly, the average human being seems to prefer a standard that he or she knows is wrong to no standard at all. After listening to the conflicting testimony on English proficiency, the judge in *Aspira of New York, Inc., et al. v. Board of Education of the City of New York, et al.*,(394 F. Supp. 1975) concluded:

> The most vivid point to emerge from all the argumentation is that we confront an enormous amount of speculation and uncertainty... (*Aspira*, 1975: 1161).

"Without approaching confidence or certainty," (p. 1164) the Court defined the plaintiff class as Hispanic students who scored at or below the 20th percentile on the English L.A.B.,[5] but higher on the Spanish L.A.B. The Court then went on to say:

> The crudity of this formulation is acknowledged on all sides. It is not possible to say with precise and certain meaning that an English-version score at

> a given percentile is similar to the same percentile score on the Spanish version...But we are merely a court, consigned to the drawing of lines, and we do the best we can (p. 1168).

Not long after the 1975 *Aspira* decision, the National Institute of Education analyzed the whole area of relative language assessment for the U.S. Department of Education and found no agreement as to what language proficiency is and general agreement that language proficiency tests are unreliable and invalid.

> ...In addition to such problems as low reliability and questionable validity and variation in theoretical underpinnings, differences in quality and quantity of items selected, and the plain fact of the incredible complexity of language, there are serious practical problems associated with assessing language proficiency on the basis of these instruments. Recent empirical studies indicate that the placement of children varies (often significantly) depending on which test is used (Spolsky, in NIE, 1981:38).

More recently, Irujo, Kramsch, Dube and Yedlin (1986) surveyed the issue of language proficiency for the Massachusetts Department of Equal Educational Opportunity. They found over 20 different definitions and concluded that language proficiency is one of the most poorly defined concepts in the field of language education. Yet, Massachusetts school districts continue to use language proficiency tests to classify students as LEP or English proficient.

The IPT, used in District A of the Hakuta, Butler, Witt study, has been found to be quite unreliable by Ramirez, Yuen and Ramey (1986). Of 573 kindergarten students classified as non-english-speaking, limited-English-speaking or fluent-English-speaking in the fall of 1984 in California, 236 had moved up one category, 238 had stayed the same, and 99 had moved down one category or more two years later in the spring of 1986. Thus, according to this test, not only has 40 percent of the sample made no progress in English over two years, but 17 percent know less English than when they began.

Similar results are found with students in higher grades. Of 232 first-graders classified LEP by the IPT in the fall of 1984, 50 percent made no progress over two years, and 13 percent knew less English than when they began. Of 123 third-graders classified LEP, 48 percent seemingly made no progress and 7 percent knew less English than when they began (Ramirez, Yuen, and Ramey, 1986).

LEP students who score low in English often score low in their native tongue because the tests also measure academic ability, not just fluency. Illustrative of this phenomenon is a study of relative language proficiency among Hispanic students in California by Duncan and De Avila (1979). A majority (54) of the 101 students classified by the Language Assessment Scales (LAS) as limited or non-proficient in English were also classified as limited or non-proficient in Spanish. Of the 96 students found to be limited or non-proficient in English, less than half (42) were considered proficient Spanish speakers according to their Spanish test score.

Moreover, language proficiency tests do not agree with each other even when they are in the same language. Ulibarri, Spencer and Rivas (1980) investigating the comparability of three oral English proficiency tests used in California (the LAS, BSM, and BINL) concluded that language classification is a function of the particular test used with each test identifying different numbers of eligible students. Studies by Gillmore and Dickerson (1979), Cervantes (1982) and Pelavin and Baker (1987) found similar results. Pelavin and Baker further found that the disagreement between tests is greatest for those students who spoke some English, in particular when a reclassification decision was being made.

Not only are the tests unreliable, but they are invalid. English proficiency tests administered to English monolingual children in experiments routinely classify large percentages of them as LEP. Berdan et al. (1982) administered the Language Measurement and Assessment Instrument (LM&AI) to Cherokee students at the request of the Cherokee Nation, which wanted to determine the need for Cherokee bilingual education. Through home interviews, Berdan et al. found that 82 percent of the Cherokee students were English monolinguals. The LM&AI, however, classified 48 percent of these monolingual English-speaking children as LEP presumably in need of instruction in Cherokee so they could improve their English. In 1984, the U.S. Department of Education had the LM&AI administered to a nationally representative sample of monolingual English-speaking school-aged children. The test classified 42 percent of them as LEP (U.S. Bureau of the Census Data, 1984).

A similar experiment in Chicago (Perlman and Rice, 1979) suggests that the problem of classifying English monolingual students as limited-English-proficient is not limited solely to low-achieving students. The Chicago Board of Education administered the Language Assessment Scales (LAS)—a test used widely throughout the U.S. and one of the approved tests in California—to students who spoke only English and were above the citywide ITBS norms in reading. Almost half of the monolingual, above average, English-speaking children were misclassified as non- or limited-English speaking. Moreover, there is a developmental trend. Seventy eight percent of the English monolingual 5-year-olds, but only 25 percent of the 14-year-olds, were classified LEP.

I am also familiar with a particular instance of misclassification in California. In 1988, the principal of an elementary school in the Berkeley Unified School District, upset over the State Department of Education's compliance review, decided not to wait for the results of the home language survey[6] before testing all new Spanish-surnamed students in her school. The

5-year-old child of a professional Hispanic family in Berkeley was administered the oral portion of the IPT in this mass testing. Although he knows no language other than English and the language of their home is English, he failed the oral proficiency test, was classified as limited-English-proficient, and assigned to the Spanish bilingual program. When the family received the notice, the mother called the school, informed it of their mistake, and was allowed to withdraw her child from the bilingual education program. But what if the mother had not been a fluent English speaker and an assertive professional who understood the mistake? There is a very good chance that this child would have been assigned to the Spanish bilingual program and taught in a language he did not know. A year later this same child, who at age 5 had been classified LEP by an oral proficiency test, was classified "gifted" on the basis of a standardized achievement test. Thus, it is possible for a gifted child to fail an oral English proficiency test and be classified LEP!

To summarize, the research evidence indicates that language proficiency tests are unreliable and invalid and there is a good deal of disagreement between the different types, particularly when the students tested speak some English. The tests fail to classify as English proficient students who are fluent in English because they cannot tell the difference between a student who does not know English and a student who does not know the answer or who refuses to answer. Moreover, all test scores, whether they are English proficiency tests or standardized achievement tests, are negatively correlated with SES. There is simply no test made that does not show that relationship.

Indeed, if we simply assume that every so-called LEP student was in fact raised in a lower socioeconomic status English monolingual family, Figure 2 indicates that we should expect about 1/2 of these English monolingual students to never attain English "proficiency." But standardized test scores are not the answer, since Figure 2 also indicates that poor students from English speaking families never achieve parity with all students on standardized achievement tests. Furthermore, there is no way to eliminate inequality in test scores since the tests are periodically renormed to produce exactly this outcome. Like a dog chasing its tail, reformers try to eliminate the normal curve, but assess their efforts with tests deliberately constructed to produce a normal curve.

How Long Do Below-Average Students Need Extra Help?

Another disagreement I have with the conclusions of Hakuta et al. is their assumption that children need to be in a special classroom or need special education services if they are below average. Hakuta, Butler, and Witt apparently believe that students are always helped by special education services, but that is not necessarily the case. It really depends on whether the problem has been accurately diagnosed and what the treatment is. If, for example, an English proficient student is incorrectly classified as LEP simply because the students scores *below* average on an *English* proficiency test, the students will undoubtedly be helped if the treatment is after-school instruction or tutoring in English and other subjects that is tailored to their needs. But this is difficult and expensive, and very few school districts in the U.S. do this.

The typical treatment for students who have been diagnosed LEP occurs during the school day so the students receive no additional instruction. The treatments are: (1) a bilingual education program with native tongue instruction if they are believed to be from a Spanish speaking family and there are enough of them to fill a classroom; (2) an ESL pullout program; or (3) a structured immersion program, that is, a self-contained classroom of LEP students taught in English at a slower pace than in the mainstream classroom.

A bilingual education program in Spanish cannot help, and probably harms, a child who does not speak Spanish. Furthermore, such inappropriate treatments do in fact occur as a result of erroneous classifications produced by English proficiency tests. For example, from 1975 to 1996 in New York City, all Hispanic students were forced to take the L.A.B. regardless of their home language and if they scored below the 40th percentile and there were enough to fill a classroom, were placed in Spanish bilingual education classrooms. In fall 1998, I visited a first-grade Spanish bilingual education class in New York City composed only of Hispanic students. During the Spanish reading period, the teacher translated most of what she said in Spanish into English because there were Hispanic students in her class who understood little or no Spanish. They had been assigned to the bilingual program, not because they did not know English, but because they had scored below the 40th percentile on the L.A.B.

In 1996, the NYC school board began to require that newly enrolled Hispanic students be from a home where a language other than English was spoken before they could take the L.A.B. The number of students classified as LEP declined by 20,000 students in New York City when this policy change was implemented. Thus, at a minimum 20,000 Hispanic students were incorrectly classified as LEP solely because they scored below the 40th percentile, and some unknown percentage of them were assigned to a Spanish bilingual education program although they did not speak Spanish. It is hard to imagine how this "special service" could help the English proficiency of these children.

While perhaps not as obvious, an ESL pullout program for a child who is fluent in English can also harm a child. ESL programs take the children out of the mainstream "grade level" classroom for an hour or more a day or a few hours a week, and place them in a small group where they learn basic grammar and concepts that are well below grade level under the assumption that they do not speak English. If the children already know what is being taught in the ESL class, but still need to learn what is being taught in the mainstream classroom during the time they are pulled out for ESL, they will be harmed by the ESL class.

Similarly, a structured immersion classroom is not a beneficial treatment for a child who is fluent in English because like ESL instruction, it is also below grade-level instruction. The teacher teaches content at a slower pace because the students are assumed to not know English. If the students already know English, they will be harmed by this slower pace. In short, special education services can in fact harm students if they do not need the slower pace. This is simple logic that is ignored by Hakuta, Butler, and Witt.

Is a Year in Structured English-Immersion Enough?

What little research there is suggests that although it could take a decade for a student to reach the highest level of English language achievement they are capable of,[7] with students who come to the U.S. at earlier grades reaching it sooner than students who enter in the later grades (Rossell, 2000), all students understand enough English sometime during the first year to be able to comprehend English instruction. I base this conclusion on research conducted in Canada and the U.S. on immersion programs, research conducted in the U.S. and Europe on newcomer centers, my conversations with LEP students in bilingual and ESL classrooms around the U.S., and my conversations with former LEP students in my classes at Boston University.

The studies of French immersion programs in Canada indicate that the English-speaking students, albeit self-selected, eager language learners, understood what the teacher said to them in French sometime during the first semester of the first year. By the end of the second year they were almost the equal of French native speakers on many tests (Genesee, 1984; Swain and Lapkin, 1982).

According to Glenn and de Jong (1996), the common European program for immigrant children is to integrate kindergarten children immediately into the mainstream classroom but to provide a "reception" class for one year for those who arrive after the usual age for beginning school. In the reception classes, the focus is on laying the foundation for enrollment in the mainstream classroom. The Europeans have no illusion that the language barrier will be overcome in a year, but they do believe that a year will provide a solid foundation for older students, and that the language barrier will only be overcome when the immigrant children enrolled in a classroom where they can interact with native speakers of the target language.

These one-year programs are also found in the U.S. under a variety of labels. McDonnell and Hill (1993) found "newcomer" schools for immigrant children in every school district they studied, including the three California school districts, San Francisco, Los Angeles, and Visalia. The length of time for students in the newcomer school was six months to a maximum of one year. McDonnell and Hill describe them as follows:

> The newcomer schools in our sample are impressive places: In their clear sense of mission, innovative curricula, professional teaching staff, and links to the larger community, they represent the kinds of schools to which all children, immigrant and native born, should have access...The newcomer schools in our sample are all self-contained programs that students attend full-time *for one or two semesters* [emphasis added], and all but the Los Angeles high school operate in physically separate locations. However, there are a variety of other newcomer models, including ones that students attend for half day and then spend the remainder of the day in mainstream classes. In contrast to the schools in our sample, in which students from across a district are transported to a single site, some districts, such as Long Beach, operate newcomer classrooms on as many as a dozen different campuses. For a description of these other program models see Chang (1990) (McDonnell and Hill, 1993, pp. 97-98).

In addition to newcomer schools, there are one-year immersion programs for kindergarten students all over California and the U.S. In Chelsea, Mass., there are one-year kindergarten immersion programs for Cambodian and Vietnamese students. In New York City there are a number of one-year kindergarten immersion programs (all of them called bilingual) for non-Hispanic LEP students, as well as entire schools for newcomers. One in particular, is the one-year kindergarten immersion program for Chinese students at the Sampson School (P.S. 160) in Brooklyn. In Boston, there is a one-year kindergarten immersion (called bilingual) program for Cape Verdean students at the Mason School. Although Mason School parents have the option of going on to a Cape Verdean "bilingual" program at another school for first grade, very few do that. The conclusion of the teachers and the parents of LEP students at this school is that one year is enough. Within one year, students comprehend enough English to be active participants in the mainstream classroom, although they have a long way to go before they reach their full capacity in English.

I have also had conversations with LEP students in public schools in California, Massachusetts, New York City, and St. Paul, Minn. In most ESL classrooms I have been in, there are one or two students who are working independently because they already know what is being taught. I have taken the opportunity to talk to these students about how long it took them before they could understand what the teacher was saying in English when they entered the school. Those who started in September, having just come from a foreign country, believe they understood what the teacher was saying by the Christmas break. I have also discussed this issue with students in my classes at Boston University who had immigrated to this country as children. None had ever been assigned to a bilingual education class, and all believed they could understand the teacher completely by the end of their first year in an English speaking classroom.

It may be that a few students would be better off staying a little longer than a year in a structured immersion classroom. We simply do not know. What we do know is that we cannot rely on test results such as those presented in the Hakuta, Butler, and Witt report to accurately place or exit students from programs because those standards will result in more than half the students never being classified English proficient, even if that is the only language they know.

This is not just hypothetical, it actually occurs. Table 3 shows the annual reclassification rates for LEP students in California from 1981-82 through 1998-99 (the first year of Proposition 227) using standards such as those in the Hakuta, Butler, Witt report. About 5 percent to 7 percent of LEP students are redesignated English proficient each year in California. If we add up these annual reclassification rates, less than half of a kindergarten cohort that began school in 1992-93 would be redesignated English proficient by the end of their elementary school career in 1998-99, although there is no way the others could not be fluent in English after this time period.

Thus, Proposition 227 is deliberately worded to limit the time period in a separate below-grade level classroom to one year, not because anyone thinks non-English speaking children will have mastered English in one year, but because what evidence there is suggests that sometime during their first year immigrant children will understand enough English so that they will be better off in a grade-level mainstream classroom than in a remedial classroom. Furthermore, if a time limit were not specified in the legislation, more than half of them would never be mainstreamed, no matter how fluent they were in English.

Table 3.

Redesignation Rates for English Learners (Limited-English-Proficient Students) and Cumulative Redesignation Rates for 1992-93 Kindergarten Cohort in California 1981-82 to 1998-99

Year	Number Of LEP Students	% of K-12 Enrollment	# of Students Redesignated FEP	% Redesignated of Previous Year's LEP	Change from Previous Year	1992 Cohort School Grade	Cumulative % Redesignated FEP w/ Assumption of Same Students in Cohort
1998-99	1,442,692	24.7%	106,288	7.6%	0.6%	6th	44%
1997-98	1,406,166	24.6%	96,545	7.0%	0.3%	5th	37%
1996-97	1,381,393	24.6%	89,144	6.7%	0.3%	4th	30%
1995-96	1,323,767	24.2%	81,733	6.5%	0.5%	3rd	23%
1994-95	1,262,982	23.6%	72,074	5.9%	0.4%	2nd	16%
1993-94	1,215,218	23.1%	63,379	5.5%	0.4%	1st	11%
1992-93	1,151,819	22.2%	54,530	5.1%	-0.6%	Kind.	
1991-92	1,078,705	21.1%	55,726	5.6%	0.0%		
1990-91	986,462	19.9%	49,001	5.7%	-1.5%		
1989-90	861,531	18.1%	53,223	7.2%	-1.2%		
1988-89	742,559	16.1%	54,482	8.4%	-1.0%		
1987-88	652,439	14.5%	57,385	9.4%	0.0%		
1986-87	613,224	14.0%	53,277	9.4%	-1.1%		
1985-86	567,564	13.3%	55,105	10.5%	0.2%		
1984-85	524,076	12.6%	50,305	10.3%	-0.1%		
1983-84	487,835	11.9%	47,503	10.4%	-1.8%		
1982-83	457,540	11.2%	52,504	12.2%	-3.0%		
1981-82	431,449	10.7%	57,336	15.2%			

Source: State Department of Education, Language Census Reports for California Schools, www.cde.ca.gov.

Endnotes

[1] Kenji Hakuta, Yuko Goto Butler, and Daria Witt, January 2000, "How Long Does It Take English Learners to Attain Proficiency?" The University of California Linguistic Minority Research Institute, Policy Report 2000.

[2] While it might seem to be common sense that a child who receives special education services will be better off than one that does not, the most common finding of the research evaluations of special education services such as Title I, Headstart, and bilingual education over the last decade is no effect.

[3] The MacMillan test is a standardized achievement test that District B uses as an English proficiency test by establishing its own criterion for "proficiency."

[4] The median is that point at which 50 percent of the cases are above and 50 percent are below.

[5] This was changed to the 40th percentile in 1989.

[6] A home language survey is the first step in identifying a new student as potentially LEP in school districts in the U.S. Typically, a new student will take an English proficiency test only if a language other than English is or was spoken by someone in the home.

[7] The highest level of English that a student is capable of is different from attaining parity with native English speakers or a test publisher's standard. Determining the highest level of English an LEP child is capable of requires a sophisticated research design that would attempt to determine an LEP child's intelligence through nonverbal tests and then the standardized test score received by native English speaking children of that intelligence level. When the LEP child has reached the test score of the native English-speaking students of their intelligence level, they are more or less at the highest level of English they are capable of.

Bibliography

Baker, Keith and Christine Rossell. 1987. "An Implementation Problem: Specifying the Target Group for Bilingual Education." *Educational Policy* 1 (2): 249-270.

Berdan, Robert, Alvin So, and Angel Sanchez. 1982. "Language Among the Cherokee, Patterns of Language Use in Northeastern Oklahoma, Part 1, The Preliminary Report." Los Alamitos, Calif.: National Center for Bilingual Research.

Cervantes, Robert A. 1982. "Entry into and Exit from Bilingual Education Programs." Washington, D.C.: E.H. White Inc.

Chang, H. N. 1990. *Newcomer Programs: Innovative Efforts To Meet the Educational Challenges of Immigrant Students.* San Francisco: California Tomorrow.

Cummins, Jim. 1981. "Immigrant Second Language Learning," *Applied Linguistics*, 11, 132-149.

Duncan, Sharon E. and Edward A. De Avila. 1979. "Relative Language Proficiency and Field Dependence/Independence." Paper presented at the annual meeting of TESOL, Boston.

Genesee, Fred. 1984. "French Immersion Programs," in Stan Shapson and Vincent D'Oyley, eds., *Bilingual and Multicultural Education: Canadian Perspectives*, Clevedon, Avon, England: Multilingual Matters.

Gillmore, G. and A. Dickerson. 1979. "The Relationship Between Instruments

Used for Identifying Children of Limited English Speaking Ability in Texas." Houston, Texas: Education Service Center [Region IV].

Glenn, Charles L. and Ester J. de Jong. 1996. *Educating Immigrant Children: Schools and Language Minorities in Twelve Nations,* New York: Garland Publishing.

Irujo, S., C. Kramsch, N.C. Dube and J. Yedlin. 1986. "Language Proficiency." Presented at Language and the World of Work in the Twenty-First Century symposium sponsored by the Bureau of Transitional Bilingual Education, Massachusetts Department of Education.

Klesmer, H. 1993. "E.S.L. Achievement Project: Development of English as a Second Language Achievement Criteria as a Function of Age and Length of Residence in Canada," Report prepared for the North York Board of Education.

McDonnell, Lorraine M. and Hill, Paul T. 1993. *Newcomers in American Schools: Meeting the Educational Needs of Immigrant Youth.* Santa Monica, Calif.: Rand Corporation.

Pelavin, Sol and Keith Baker. 1987. "Improved Methods of Identifying Who Needs Bilingual Education." Paper prepared for delivery at the annual meeting of the American Educational Research Association, Washington, D.C.

Perlman, C. and W. Rice Jr. 1979. "A Normative Study of a Test of English Language Proficiency." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, Calif.

Ramirez, J. David, Sandra D. Yuen and Dena Ramey. 1986. "Second Year Report: Study of Immersion Programs for Language Minority Children." Alexandria, Va.: SRA Technologies, Inc.

Rossell, Christine. 2000. "Teaching Language Minorities; Theory and Reality," in *City Schools: Lessons from New York,* Diane Ravitch and Joseph Viteritti, eds., Baltimore: The Johns Hopkins University Press.

Rossell, Christine and Keith Baker. 1988. "Selecting and Exiting Students in Bilingual Educational Programs." *Journal of Law and Education* 17 (4): 589-623.

Spolsky, Barnard. 1978. *Educational Linguistics.* Rowley, Mass.: Newbury House Publishers cited in National Institute of Education. 1981. "Report on the Testing and Assessment Implications of the Title VI Language Minority Proposed Rules." Washington, D.C.: National Institute of Education.

Swain, Merrill and Sharon Lapkin. 1982. *Evaluating Bilingual Education: A Canadian Case Study.* Clevedon, Avon, England; Multilingual Matters.

Ulibarri, D., M. Spencer, and G. Rivas. 1980. Comparability of Three Oral Language Proficiency Instruments and Their Relationship to Achievement Variables." Report submitted to the California State Department of Education.

U.S. Bureau of the Census, Data for the Office of Planning, Budget and Evaluation, Decision Resources. 1984. "1984 Analysis." Washington, D.C.: Department of Education.

Wright, E. N. and Ramsey, C. 1970. "Students of Non-Canadian Origin: Age on Arrival, Academic Achievement and Ability." Research Report #88, Toronto Board of Education.

About the Authors

RALPH E. BEALS, Ph.D., is Clarence Francis professor of economics at Amherst College, where he has served as Acting Dean of the Faculty and Acting President. Among other publications, he is the author of *Economic Statistics*, an econometrics text that has been widely used in public policy programs.

ROGER CLEGG is vice-president and general counsel of the Center for Equal Opportunity in Washington, D.C., where he writes, speaks and conducts research on civil rights issues. A graduate of Rice University and Yale Law School, Mr. Clegg served for 11 years at the U.S. Justice Department, including four years as a deputy in the civil rights division.

WILLIAM A. MEHRENS is professor of educational measurement at Michigan State University. His specializations are the legal issues and the consequences of large-scale assessments. He has published widely in the field, and is a past president of the National Council on Measurement in Education.

SUSAN E. PHILLIPS combines the unique educational background of a doctorate in educational measurement and a law degree. She has published and lectured extensively on legal issues in assessment. As a consultant to many state departments of education, Dr. Phillips has focused on the psychometric and legal issues related to the defensibility of high-stakes assessments.

ROSALIE PEDALINO PORTER is the author of *Forked Tongue: The Politics of Bilingual Education* (Basic Books, 1990; *2nd* edition Transaction Publishers, 1995) and is the editor of *READ Perspectives*, a refereed journal published by the Institute for Research in English Acquisition and Development (READ) in Washington, D.C. She has lectured on education policy for language minority students on behalf of the U.S. State Department in Bulgaria, Finland, Israel, Italy, Japan and Turkey.

CHRISTINE H. ROSSELL is professor of political science at Boston University. She has published four books and more than 50 articles on school desegregation, bilingual education and education policy. Her most recent book, *Bilingual Education in Massachusetts: The Emperor Has No Clothes,* was published in 1996 by Pioneer Institute of Boston.

ROBERT E. ROSSIER is an independent consultant on English language programs for immigrant students for several California school districts. He has lectured and written on various aspects of bilingual education policy and pedagogy, and served on the National Advisory and Coordinating Council on Bilingual Education in Washington, D.C.

READ
PERSPECTIVES

is now an annual, serial publication.
Each issue will focus on a theme of importance in the education of limited-English students in U. S. public schools.

CALL FOR MANUSCRIPTS

Volume VIII, Fall 2001, will focus on the theme
Accountability for Bilingual Students: Focus on California

Articles on student achievement in schools and districts; Academic Performance Index results; promising new program initiatives; effects of Proposition 227 on students' second language learning and academic progress.

Deadline for manuscripts: March 15, 2001
Length of articles and honoraria are negotiable

Send proposals to: The Editor, *READ Perspectives*
815 15th Street, N.W., Suite 928
Washington, DC 20005

CEO
READ Institute
Center for Equal Opportunity
815 15th Street, N.W.,
Suite 928
Washington, DC 20005
Phone (202) 639-0803
Fax (202) 639-0827
http://www.ceousa.org